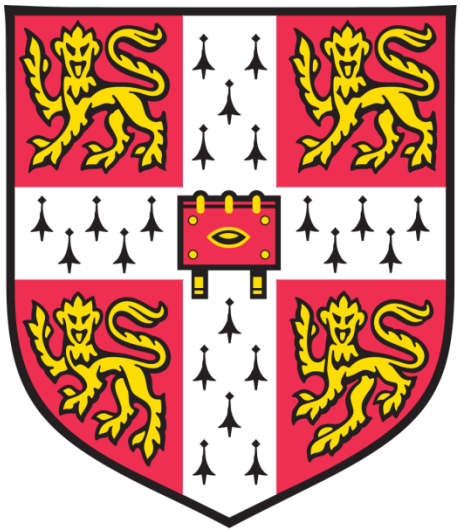
Investigating the Non-globular Proteins of the Canonical Wnt Signalling Pathway



Benjamin Martin Smith

This dissertation is submitted for the degree of Doctor of Philosophy

Corpus Christi College, University of Cambridge

August 2017

Declaration

This dissertation contains my own work undertaken in the Department of Pharmacology, University of Cambridge between October 2013 and August 2017. Any work undertaken by, or in collaboration with others is specifically indicated in the text.

A small proportion of the work described in Chapter 3 has been previously submitted for the Certificate of Postgraduate Study (CPGS) in June 2014. Apart from this, this dissertation does not contain materials that have previously, or, is being concurrently submitted for any other degree, diploma or qualification at the University of Cambridge or any other University or similar institution.

This dissertation does not exceed 60,000 words.

Benjamin Martin Smith

August 2017

Acknowledgements

First and foremost I would like to thank my supervisor, Prof. Laura Itzhaki, for giving me the opportunity to do this PhD and for all of her advice and guidance throughout. I would also like to thank Drs. Pam Rowling, Ana Chattopadhyay, Elin Sivertsson and Ben Murton for teaching me all the techniques required to transform this undergraduate chemist into a real biophysicist. I am also grateful to all the members of the Itzhaki group for making this an amazing place to work for these last four years. Albert, Alex, Aurora, Ewan, Grasilda, Marie, Owen, Rohan, Sarah, Sonja, Tonia, Wenshu, Yalda and Zhen – thank you for all the great conversation and great cake over the years.

I am very grateful to Prof. Chris Dobson for his generosity in funding this PhD, as well as the additional financial support I have received from the Department of Pharmacology and from Corpus Christi College to allow me to attend conferences.

I would also like to thank all the members of the Department of Pharmacology for being so welcoming. I am grateful to Jon, Paul and Barney for always being there to fix things when they break, to Lesley, Lisa and Roger for encouraging me to engage in teaching and outreach, and to my mentor, Dr. Matthew Harper, his great advice and entertaining conversations.

Life outside the lab also has a huge influence on life within it, so I would like to give a big thank you to all my friends outside the lab. In particular, I would like to thank Watsy, who has introduced me to many of my closest friends, as well as stimulating my imagination and creativity; two often overlooked skills in science. Thanks you for the years spent building character, learning how to play the hand I have been dealt, and rolling with the best of them.

Last but certainly not least my mum. Thank you for all the love and support for these last 26 years. Thank you for lifting me back up when I am down and being the first to call me out for being a tool. Thank you for instilling in me the burning curiosity that has driven me to this point.

Abstract

The canonical Wnt pathway is a vitally important signalling pathway that plays an important role in cell proliferation, differentiation and fate decisions in embryonic development and in the maintenance of adult tissues. The twelve Armadillo (ARM) repeat-containing protein β-catenin acts as the signal transducer in this pathway and is continuously degraded in the cytosol by the β-catenin destruction complex (BDC). Upon receiving the Wnt signal the BDC is inactivated, allowing β-catenin to accumulate in the cytosol and be transported to the nucleus where it binds to the TCF/LEF family of transcription factors, inducing the expression of cell cycle promotor genes.

In this Thesis I describe investigations into the roles of leucine-rich repeat kinase 2 (LRRK2) and the transcription factor TCF7L2 within this signalling pathway. LRRK2 is a large multi-domain protein with strong links to Parkinson's disease and suggested to play a role in inactivating the BDC in response to the Wnt signal. A recent paper proposed that the previously uncharacterised regions of LRRK2 contain a series of tandem repeat sub-domains. I began an investigation into these sub-domains but I was unable to produce soluble protein constructs despite the use of a range of common techniques, and so I was forced to conclude this project early.

The main body of this thesis focuses on the interaction between the intrinsically disordered TCF7L2 and the repeat protein β-catenin, a very long interface of approximately 4800 $Å^2$ that spans from the third to the eleventh ARM repeat of β-catenin and residues 12 to 50 of TCF7L2, as determined by X-ray crystal structures. First, a fluorescence reporter system for the binding interaction was developed and used to determine the kinetic rate constants for the association and dissociation of the wild-type construct using stopped-flow fluorescence spectroscopy and time-dependent fluorescence spectroscopy. It was found that association of TCF7L2 and β-catenin was rapid (7.3 ± 0.1 $x10^7$ $M^{-1}s^{-1}$) with only a single phase was observed, whereas dissociation was biphasic and slow (5.7 ± 0.4 $x10^{-4}$ s^{-1}, 15.2 ± 2.8 $x10^{-4}$ s^{-1}). Using either of these two dissociation rate constants the calculated K_d value obtained is much lower than the values previously reported in the literature (8 ± 1 / 20 ± 2 pM compared with 16 nM).

This reporter system was then used to investigate the striking variability between three crystal structures previously obtained for the TCF7L2-β-catenin complex, in which different regions of TCF7L2 show different elements of secondary structure. Mutational analysis revealed that the interface residues on TCF7L2 identified in these structures make little or no contribution to the overall binding affinity, pointing to a transient nature of these contact in solution and suggesting that the observed differences between the structures are due to differences in crystal packing.

Further experiments into the effect of osmolarity on the binding equilibrium and kinetics supported this conclusion and suggest a change in the association/dissociation mechanism as a function of ionic strength.

Lastly, further mutational analysis of TCF7L2 revealed two regions that contribute particularly strongly to the binding kinetics, suggesting that TCF7L2-β-catenin assembly proceeds via a two-site avidity mechanism. Some of the most destabilising variants display two additional dissociation phases, indicating the presence of an alternative dissociation pathway that is inaccessible to the wild-type.

In summary, the results presented here provide insights into the kinetics of molecular recognition of a long intrinsically disordered region with an extended repeat protein surface, a process shown to involve multiple routes with multiple steps in each.

Abbreviations

A_x	absorbency at x nm wavelength
aa	amino acid
ABA	abscisic acid
ABAR	putative ABA receptor
ACTR	activator for thyroid hormone and retinoid receptor
AD	subdomain of p53
Aha	azidohomoalanine
AMD	age-related macular degeneration
ANK	ankyrin
amp	ampicillin
APC	adenomatous polyposis coli
ARM	armadillo
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BDC	β-catenin destruction complex
cAMP	cyclic adenosine monophosphate
CBP	CREB binding protein
CD	circular dichroism
CDK	cyclin-dependent kinase
CK1α/γ	casein kinase 1α/γ
COR	C-terminal of ROC
CREB	cAMP response element-binding protein
CRISPR	clustered regularly interspaced short palindromic repeats
CuAAC	copper-catalysed azide-alkyne cycloaddition
d.p.	decimal place
DsbC	disulfide bond isomerase C
DMSO	dimethylsulfoxide
DTT	1,4-dithiothreitol
ε_x	extinction coefficient at x nm wavelength
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraactetic acid
ELISA	enzyme-linked immunosorbent assay

FG Nup	phenylalanine-glycine nucleoporin
FIH	factor-inhibiting hypoxia-inducible factor
FITC	fluorescein 4-isothiocyanate
Fz	frizzled
GSK3β	glycogen synthase kinase 3β
GST	glutathione S-transferase
HEAT	huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast kinase, TOR1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HDX-MS	hydrogen-deuterium exchange mass spectroscopy
HIF	hypoxia inducible factor
HPV E7	human papillomavirus E7
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IPTG	isopropyl β-D-1 thiogalactopyranoside
kan	kanamycin
LRP	LDL (low-density lipoprotein) receptor-related protein
LRR	leucine-rich repeat
LRRK2	leucine-rich repeat kinase 2
MAPKKK	mitogen-activated protein kinase kinase kinase
Mcl-1	induced myeloid leukemia cell differentiation protein
MCS	multiple cloning site
MES	2-(N-morpholino)ethanesulfonic acid
MoRFs	molecular recognition features
MPP^+	1-methyl-4-phenylpyridinium
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
M_w	molecular weight
MWC	Monod-Wyman-Changeux
NADPH	reduced nicotinamide adenine dinucleotide phosphate
NCBD	nuclear co-activator binding domain
NCL	native chemical ligation
NDSB	non-detergent sulfobetaine
NMR	nuclear magnetic resonance
NPC	nuclear pore complex

NTA	nitrilotriacetic acid
OD_{600}	optical density at 600 nm wavelength
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	Parkinson's disease
PDB	Protein Data Bank
PEG	polyethylene glycol
PMT	photomultiplier tube
PNK	polynucleotide kinase
PP2A	protein phosphatase 2A
PTM	post-transcriptional modificiation
PUMA	p53 upregulated modulator of apoptosis
ras	rat sarcoma gene family
Rb	retinoblastoma tumor suppressor
RCM	ring-closing metathesis
RTH	round-the-horn
ROC	ras of complex proteins
RVD	repeat-variable diresidue
SAX	small angle X-ray scattering
SDM	site-directed mutagenesis
SDS	sodium dodecylsulfate
SEC	size exclusion chromatography
SOAR1	suppressor of the ABAR-overexpressor 1
SOC	super optimised broth with catabolite repression
SPPS	solid-phase peptide synthesis
θ_x	ellipticity at x nm wavelength
TAZ	transcriptional adapter zinc finger
TALE	transcription activator-like effector
TALEN	TALE nuclease
TCF7L2	transcription factor 7-like 2
TFE	2,2,2-trifluoroethanol
TLE	transducing-like enhancer of split
T_m	melting temperature

TMAO	trimethylamine N-oxide
Tris	tris(hydroxymethyl)aminomethane
VEGF	vascular endothelial growth factor
WD40	tryptophan-aspartic acid-40

Contents

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Tandem Repeat Proteins

Repeat protein is the term used to describe any protein that contains some repeated element of sequence and/or structure, with the current estimate being that 14% of all proteins contain segments of duplicated sequence (Marcotte *et al.*, 1999). Generally the more complex an organism, the greater proportion of their genome will encode for proteins containing repeating structure, with 25% of human proteins containing repeat motifs (Marcotte *et al.*, 1999). This is likely a result of gene duplication, and the subsequent diversification of function, as a facilitator of evolution. This also explains why eukaryotic repeat proteins tend to be larger, containing more repeats, than those seen in prokaryotes (Heringa, 1998).

More often than not, discussion of repeat proteins refers to tandem repeat proteins i.e. proteins containing tandem repeats of a small structural motif that stack together to form elongated structures which are distinct from "typical" globular proteins (Javadi and Itzhaki, 2013). Typically tandem repeats form highly symmetrical linear arrays, which are either structurally open and superhelical, referred to as solenoids or superhelices (Kobe and Kajava, 2000), or circularly closed (Blaber and Lee, 2012).

1.1.1 Tandem Repeat Families

There are currently twenty different classes of repeat protein (Kobe and Kajava, 2000), but this study will focus on four families in particular: three linear repeat families; Ankyrin (ANK), Armadillo (ARM) and Leucine-Rich Repeat (LRR), and circularly closed WD40 repeat family. The properties of these four families are summarised in Table 1-1.

The ankyrin (ANK) repeat is one of the most common structure repeat motifs seen in nature (Mosavi, Minor and Peng, 2002). First identified in the signalling proteins *Saccharomyces* Cdc10 and *Drosophila* Notch (Breeden and Nasmyth, 1987), the family is named for the human protein, ankyrin (Lux, John and Bennett, 1990). The motif comprises two antiparallel α-helices, linked by a β-turn, with an unstructured loop of variable length connecting adjoining repeats which stack to form a nearly linear structure with little curvature (see Figure 1-1a). Typically only four to six repeats are seen in nature making the eponymous ankyrin protein somewhat unusual, containing 24 of these repeats (Mosavi *et al.*, 2004).

Table 1-1: Characteristics of repeat protein families

Family	Amino acids per repeat	Structural motif	Solenoid structure	Consensus sequence*	Number of repeats seen in natural proteins
ANK	30-34[1]	helix-turn-helix	Open	GxTPLHlAarxGhlevvk lLLxxGAdvnakdkn[2]	4-34
ARM	40-42[3]	helix-helix-helix	Open	xxxxxVxxxxGxLPxLVxLxLx sxxdxxVxxxAaxaLsNLax[4]	6-14
LRR	20-30[5]	strand-turn-helix	Open or closed	LxxLxLxxNxLxx…xLxxLxx[6]	Up to 28
WD40	40[7]	β-propeller[8]	Closed	xvxsvxldgxxlasgsxDxti kv**WD**xxxGxClxtlxgH[8]	4-16 (typically 7)

*Capitalisation indicates residue is highly conserved, lower case indicates some degree of conservation, x indicates no significant conservation and … indicates variable repeat length.

[1] (Lux, John and Bennett, 1990); [2] (Mosavi, Minor and Peng, 2002); [3] (Peifer, Berg and Reynolds, 1994); [4] (Andrade *et al.*, 2001); [5] (Kobe and Kajava, 2001); [6] (Bella *et al.*, 2008); [7] (Neer *et al.*, 1994); [8] (Li and Roberts, 2001)

The armadillo (ARM) repeat is another ubiquitous repeat family that has been observed in all eukaryotic domains of life (Peifer, Berg and Reynolds, 1994). The family's name derives from the gene name for the *Drosophila* analogue of the human protein, β-catenin, which is named armadillo after the appearance of the *Drosophila* embryos when this gene is mutated (Coates, 2003). The ARM repeat comprises three α-helices arranged in a triangular pattern, with the third helix resting on top of the first causing a twist which results in its iconic superhelix of helices structure (see Figure 1-1b) (Huber, Nelson and Weis, 1997).

The leucine-rich repeat (LRR) motif is named for its unusually high leucine content (Takahashi, Takahashi and Putnam, 1985). This repeat comprises a β-strand-turn-helix motif, where the type of helix can vary between proteins. Most commonly it is an α-helix, although 3_{10}-helices and π-helices have been observed (Enkhbayar *et al.*, 2003). These repeats stack together to form an α/β horseshoe fold, with the β-strands coming together to form a parallel β-sheet on the concave surface and the helices on the convex face (see Figure 1-1c), the type and pattern of the helices determining the curvature of the solenoid structure (Kobe and Deisenhofer, 1994). Although the LRR proteins seen in nature do not contain enough repeats for this to be observed, given enough repeats, the structure could form an open superhelix or be circularly closed (similar to β-barrel proteins) and even the hypothetical β-Möbius Strip (Enkhbayar *et al.*, 2003).

The WD40 repeat is different from the other three as it always forms a circularly closed structure. It is also distinguished by having a generally lower level of conservation in its primary consensus sequence with the exception of its namesake tryptophan-aspartic acid dipeptide motif (Neer *et al.*,

1994; Li and Roberts, 2001). The motif comprises a four-stranded β-propeller, where four β-strands form an anti-parallel β-sheet which is twisted such that the first and fourth strand lie almost perpendicular to each other. Typically four to eight of these β-propellers come together to form a circularly closed structure (the most common being the seven-bladed β-propeller), with first strand forming a parallel β-barrel on the concave face. It is also common for the first (or last) repeat to be non-contiguously formed, with β-strands being contributed from both the N- and C-terminal parts of the domain, as demonstrated in Figure 1-1d.

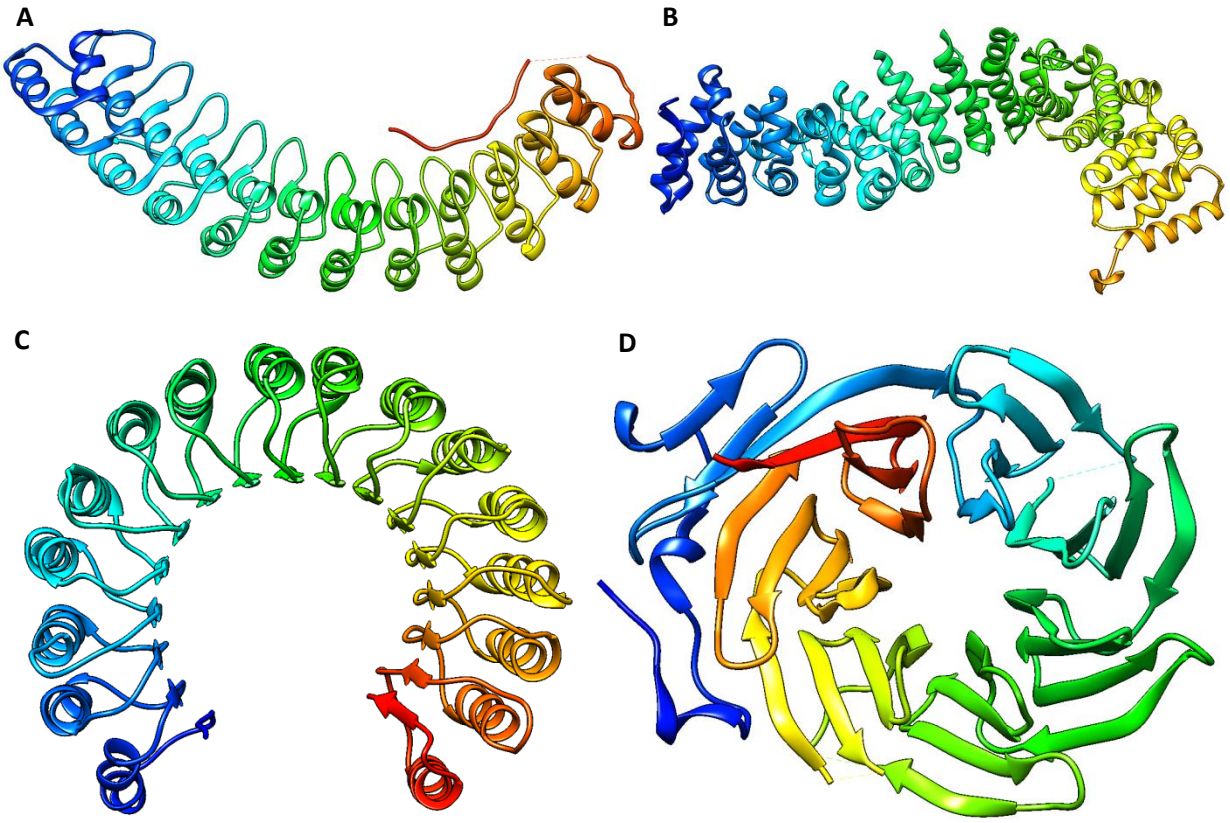


Figure 1-1 – The solenoid structures of tandem repeat proteins. Examples of proteins from the four repeat protein families discuss in this thesis: (A) the ANK subdomain of ankyrinR; (B) the ARM repeat region of β-catenin; (C) porcine ribonuclease inhibitor; and (D) the C-terminal WD40 domain of Tup1. Structures (A), (B), (C) and (D) were visualised in UCSF Chimera (Pettersen *et al.*, 2004) using the PDB IDs 1N11 (Michaely, 2002), 2Z6H (Xing *et al.*, 2008), 2BNH (Kobe and Deisenhofer, 1996) and 1ERJ (Sprague, 2000) respectively. The protein sequences are coloured blue to red from the N- to the C-terminus.

1.1.2 Biophysical Properties of Tandem Repeat Proteins

Tandem repeat proteins are characterised by their simple architecture, which results in them forming fewer long-range interactions and more short-range interactions. For linear tandem repeat proteins, this often has the consequence that individual repeats tend not to be inherently stable and are dependent on their neighbouring repeats for stability (Mello and Barrick, 2004). This property actually makes them more amenable to biophysical studies of protein folding and unfolding.

For globular proteins, folding processes typically involve the whole structure moving through a diffuse transition state structure, in a concerted two-state mechanism (Fersht, 1997). Due to their modular nature, the folding of repeat proteins tends to be less concerted and more polarised, where it has been demonstrated that a subset of repeats are completely folded in the transition state whereas the rest are completely unfolded (Tang, Fersht and Itzhaki, 2003; Bradley and Barrick, 2006; Lowe and Itzhaki, 2007; Courtemanche and Barrick, 2008).

Although in some proteins, individual repeats are capable of folding independently (like "beads-on-a-string" (Ferreiro and Komives, 2007)), most fold with a high degree of cooperativity, being dependent on the inter-repeat interactions to provide stability. In the case of smaller repeat proteins (<150 amino acids), such as the four ankyrin repeat protein, myotrophin (Lowe and Itzhaki, 2007), folding resembles the concerted two-state folding seen for globular proteins, with only the folded and completely unfolded states populated at equilibrium (Ferreiro and Komives, 2007).

Larger repeat proteins tend to have more complex folding mechanisms and typically do not exhibit two-state behaviour (Werbeck and Itzhaki, 2007; Javadi and Main, 2009). Cooperativity between repeats often results in the formation of stable subdomains leading to the population of intermediates or the potential for alternative parallel folding and unfolding pathways (Tang, Fersht and Itzhaki, 2003; Bradley and Barrick, 2006; Werbeck and Itzhaki, 2007; Tsytlonok *et al.*, 2013).

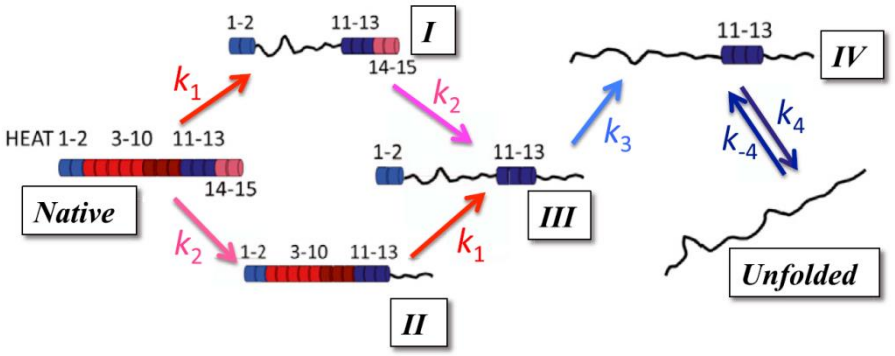


Figure 1-2 – Parallel unfolding pathway in PR65/A. The proposed multistep unfolding pathway of the giant 590 residue, 15 HEAT repeat protein, PR65/A, involving a parallel pathway where repeats 3 to 10 and repeats 14 and 15 can unfold in independent events in either order. Reproduced from Tsytlonok *et al*. (2013).

Due to their dependence on long-range contacts for stability, it is very difficult, even impossible in many cases, to dissect the contribution of individual parts of a globular protein give to its stability (Kloss, Courtemanche and Barrick, 2008). Conversely the modular nature of tandem repeat proteins makes it possible to deconvolute the contribution of individual repeats to overall protein stability by breaking them down into their smaller constituent parts (Rowling *et al.*, 2015). The traditional technique of destabilisation of the folding intermediate can also be used. However this perturbation can potentially result in the population of alternative intermediates and rerouting the folding

pathway, as has been observed for a number of ANK repeat proteins (Lowe and Itzhaki, 2007; Tripp and Barrick, 2008; Werbeck *et al.*, 2008; Tsytlonok *et al.*, 2013; Hutton *et al.*, 2015), as was predicted by *in silico* methods (Ferreiro *et al.*, 2005).

1.1.3 Natural and Artificial Functions of Tandem Repeat Proteins

The formation of the stable extended structures seen in linear tandem repeat proteins provides nature with a number of selective advantages over globular proteins. The primary advantage is that this provides a larger surface area for protein-partner interactions (Grove, Cortajarena and Regan, 2008). And because fewer residues are dedicated to protein topology, a greater degree of sequence variation is permissible within repeat protein families than in globular protein families, allowing a greater variety of unique binding surfaces to be presented for a given structural motif (see Table 1-2). A larger surface area also allows for increased promiscuity, where a single protein is capable of binding multiple binding partners allowing them to serve multiple functions, e.g. LRRK2 (see Section 1.4) or β-catenin (see Section 1.5), or act as hub proteins allowing cross-talk between different signalling pathways, e.g. SOAR1, which is thought to enable crosstalk between the PYR/PYL/RCAR- and CHLH/ABAR-mediated signalling pathways in plants (Jiang *et al.*, 2014).

Extended surface area is not the only way in which the unique properties of tandem repeat proteins can provide additional functionality. The inherent instability of the internal repeats can provide the flexibility necessary for molecular recognition. For example, IκBα contains two weakly folded ANK repeats that are essential for its ability to cause NF-κB to dissociate from DNA. Bergqvist *et al.* (2009) were able to demonstrate that stabilising these two repeats actually impaired this function. A similar thing was show for the HEAT-repeat protein, PR65/A binding to protein phosphatase 2A by Tsytlonok *et al.* (2013).

Alternatively, partial unfolding or transient unfolding events as a result of low stability can be essential in facilitating post-transcriptional modifications (PTMs) to enable further functionality. Löw *et al.* (2009) showed that phosphorylation of the ANK repeat-containing tumour suppressor protein, $p19^{INK4d}$, at S^{76} destabilises repeats 1 and 2 causing them to unfold and prevent binding to CDK4 and CDK6, acting as a molecular switch. This also has the effect of exposing K^{62} for ubiquitination and subsequent degradation, providing a regulatory mechanism. Another example would be asparagine β-hydroxylation of ANK repeats by FIH, which has been shown to stabilise ANK repeat domains (Kelly *et al.*, 2009), including Notch (Coleman *et al.*, 2007), Tankyrase-2 (Cockman *et al.*, 2009) and IκBα (Devries *et al.*, 2010). The data suggests that partial or complete unfolding of the domains is required for this process (Barrick, 2009) which would explain why Kelly *et al.* did not observe asparagine β-hydroxylation for all ANK repeat domain-containing proteins.

Table 1-2: Functions of tandem repeat proteins found in nature

Family	Biological Function	Proteins
ANK[1]	Transcription Factor	Mbp1 (late G1 phase)
		Swi4 & 6 (late G1 phase)
	Oncoprotein / Transcriptional Regulator	BCL-3
		GABPβ
		Gankyrin
		IκB
		NPR1
	Endocytosis	ANKRA
	Regulation of muscle stress respeonse	Ankrd2
		CARP
		DARP
	Polyadenylation Inhibition	BARD1
	Chloroplast signal recognition particle	cpSRP43
	Cell-cell adhesion	Integrin linked kinase
	Tumor suppressor / cell cycle regulator	INK4 family ($p15^{INK4b}$, $p16^{INK4a}$, $p18^{INK4c}$ & $p19^{INK4d}$)
		53BP2
	Development / Cell-fate determination	Myotrophin/V1
		Notch
	Cytoskeletal assembly and other structural functions	MYPT1 (myosin phosphorylation)
		P85 (actin regulation)
		SHANK1 & 2 (postsynaptic density scaffolding)
		Ankyrin
	Poly(ADP-ribose) polymerase	Tankyrase
	Signal transduction	Tvl-1
ARM[2]	Nuclear transport	Importin family
	Cell-cell adhesion	α-catenin
		β-catenin
		δ-catenin (p120)
	Wnt signalling pathway	armadillo/β-catenin
		adenomatous polyposis coli (APC)
	Cytoskeletal assembly and other structural functions	PF16/SPAG6 (sperm flagellum motility)
	Vesicular trafficking	vac8p (transport to vacuole)
	RNA regulation	Smg-GDS (mRNA quality control)
	Muscle regulation	SMAP/KAP
	Light sensing and response (plants)	PHOR1
	Gibberellin signalling	ARC1
	Exocytosis	Nd9p (membrane fusion)
LRR[3]	RNA regulation	RNase Inhibitor
		RNA1 (*S. cerevisiae*)
	RNA processing	U2 snRNP A' (splicing)
	Extracellular matrix binding	Biglycan
		Decorin
		Fibromodulin
	Corneal transparency	Lumican
	Bone morphogenesis	Osteoinductive factor
	Cell-cell adhesion	Platelet GP Ibα & V

Table 1-2 (cont.)

Family	Biological Function	Protein
LRR[3] (cont.)	Bacterial virulence / invasiveness	YopM (*Y. pestis* - bubonic plague)
		Internalin (*L. monocytogenes* – listeriosis)
	Development / Cell-fate determination	Toll (embryo development)
		Chaoptin (photoreceptor-cell development)
	Cytoskeletal assembly and other structural functions	Slit (axon development)
		Connectin (synapse development)
		Oligodendrocyte myelin GP (myelination)
	Immune response	CD14 (bacterial lipopolysaccharide receptor)
	Receptor protein kinase	Trk, TrkB & TrkC
		TMK1
	Signal transduction	LH-DG receptor
		FSH receptor
		TSH receptor
		Adenylate cyclase
		GRR1
	DNA repair	RAD1 & 7
	DNA recombination	DRT100
	Oncoprotein / Transcriptional Regulator	CCR4
	Mitosis	sds22
	Reticulated membrane	p34 ribosome-binding protein
	Inflammation response	Carboxypeptidase N
WD40[4]	Signal transduction	Gβ protein
		RACK1
		Stratin
		STE4
		LIS1
		MSI1
		PR55
		PLAP RbAp-48
	RNA synthesis	TATA box-binding protein associated factors
	RNA processing	Tup1
		CstF
	Chromatin assembly	CAF-1
		HIR1 & 2
	Vesicular trafficking	coatomer
		COPα & β
		SEC13
	Cytoskeletal assembly and other structural functions	MAP
		dynein
		Arp2 & 3
	Cell cycle regulation	CDC, CDC4, CDC20 & CDC40
		Coronin
		Mad2
	Apoptosis	Apaf-1
		Dark, CED-4

[1] (Mosavi *et al.*, 2004), [2] (Coates, 2003), [3] (Kobe and Deisenhofer, 1994), [4] (Li and Roberts, 2001)

The properties of tandem repeat proteins also lend themselves well for use as building blocks in the rational design of artificial proteins with novel functionality (Javadi and Itzhaki, 2013). Their simple, modular topology is unreliant on long range interactions for stability, instead depending on a few key residues for stability.

Consensus sequences for these repeat protein families (such as those given in Table 1-1) are the results of millions of years of selective pressure on natural proteins and only through careful study of the different sequences of proteins within a family are we able to find them (Forrer *et al.*, 2004). From the consensus sequence we have been able to identify the residues that are critical to stability and morphology and those that are important to molecular recognition and functionality. This has allowed us to modify natural proteins to improve stability without compromising function e.g. IκBα (Ferreiro and Komives, 2007).

Similarly, it is possible to modify function without compromising stability, a prime example of this being the development of Transcription Activator-Like Effector Nucleases (TALENs) (reviewed (Sun and Zhao, 2013)). TALEs are a repeat protein family secreted by the bacterial plant pathogens of the genus *Xanthomonas*. Each 34 amino acid repeat contains a pair of residues, known as the repeat-variable diresidue (RVD) which can recognise and bind to a specific base pair in dsDNA in a one-to-one manner. Li *et al.* (2011) fused TALEs to a FokI DNA-cleavage domain to create TALENs, a gene editing tool akin to CRISPR/Cas9 or zinc-finger nucleases (reviewed by Yin, Kauffman and Anderson (2017))

Also from these consensus sequences, we have been able to create consensus-designed repeat sequences to use as building blocks in the design of artificial proteins with novel functionality. These consensus design proteins, created from near identical repeats, often demonstrate stability far higher than are seen for most globular and natural repeat proteins (Kohl *et al.*, 2003) due to their highly optimised inter-repeat interactions. For example, full-consensus designed ankyrin repeat proteins (DARPins) of sufficient length cannot be fully unfolded, even at high temperatures in the presence of high concentrations of denaturants (Wetzel *et al.*, 2008). However it is often the case that the stability is tuned by changing some of the consensus residues.

These highly stable building blocks have been used to create novel proteins for various purposes (reviewed by Plückthun (2015)). Novel DARPins have been created for use in circumventing viral infection (Veesler *et al.*, 2009), targeting breast cancer tumours (Theurillat *et al.*, 2010) and there are currently DARPins in clinical trial for the treatment of AMD (Souied *et al.*, 2014) (Phase III at time of writing) and targeting VEGF (Pre-clinical) (see www.molecularpartners.com).

The ANK repeat is not the only repeat motif for which a consensus designed repeat has been created and developed: consensus-designed tetratricopeptide repeat proteins (cTPRs) have been used to create functionalised smart gels and films (Grove *et al.*, 2012 & 2013); and designed ARM repeat proteins are emerging as an effective alternative to monoclonal antibodies for peptide binding applications such as western blots (Varadamsetty *et al.*, 2012).

1.2 Intrinsically Disordered Proteins (IDPs)

Intrinsically disordered proteins (IDPs) are a class of proteins which, as the name implies, are lacking in fixed tertiary (and often secondary) structure. In a field formally dominated by the structure-function paradigm (born from the "lock-and-key" (Fischer, 1894) and later "induced fit" (Koshland, 1958) models of enzyme catalysis), proteins that functioned without structure were seen as rare and obscure exceptions to the norm. It was not until the turn of the millennium that the hypothesis that such naturally flexible species could represent a new and unique class of proteins was postulated, as a result of being highlighted by a number of seminal papers (Wright and Dyson, 1999; Uversky, Gillespie and Fink, 2000; Dunker *et al.*, 2001; Tompa, 2002).

The structure-function paradigm represented a major barrier to this discovery. While the existence of proteins lacking in secondary structure has been known for many decades (Linderstrøm-Lang and Schellman, 1954), it was always assumed that they were an artefact of the experimental conditions and not the deliberate design of nature. Consequently, "natively denatured" proteins (Schweers *et al.*, 1994) were rediscovered and renamed multiple times over the intervening years (Uversky (2011) lists 19 different terms used), making it difficult for past researchers to compare findings. Since then the terms IDP and IDR (intrinsically disordered region, denoting regions within larger, mostly structured proteins that have no defined structure) have become the naming convention, having established that function can arise from disorder.

Additionally the properties of IDPs (discussed in Section 1.2.1) mean that many of the biophysical techniques used to characterise structured proteins either fail to report or misreport the properties of IDPs. IDPs are inherently more sensitive to degradation by proteases and so are underrepresented when purifying from cell homogenates, which are protease rich environments (Dyson and Wright, 2005). Their disorder also either makes them unresolvable in crystal structures or often prevents them from crystallising in the first place. When the field of protein science is fixated on acquiring crystal structures in order to elucidate function, those that could not be seen were ignored. However, over the last two decades since the idea of functional disorder was established, a number of new techniques have been developed, as well as advancements in classical techniques, to help

identify and characterise this fascinating class of proteins. Techniques like nuclear magnetic resonance (NMR), small angle X-ray scattering (SAXS), circular dichroism (CD) and single molecule techniques have provided some insight into the partial or transient structures formed by IDPs (Tompa, 2011 & 2012; Gibbs and Showalter, 2015; Kumagai, DeMarco and Lopes, 2017).

This new wave of structural data pertaining to IDPs has been collated into a database called DisProt (Sickmeier *et al.*, 2007; Piovesan *et al.*, 2017), an equivalent of the PDB (Berman *et al.*, 2000; Rose *et al.*, 2017) for the "unfoldome" (Uversky, 2010) i.e. proteins containing intrinsic disorder. From data repositories like DisProt, it has been possible to develop software tools such as PONDR-FIT (Xue *et al.*, 2010) and MetaDisorder (Kozlowski and Bujnicki, 2012) to accurately predict regions of protein disorder, and with the aid of these tools it has been estimated that 10-35% of prokaryotic proteins and 15-45% of eukaryotic proteins contain disordered regions (>30 residues) (Tompa, 2012).

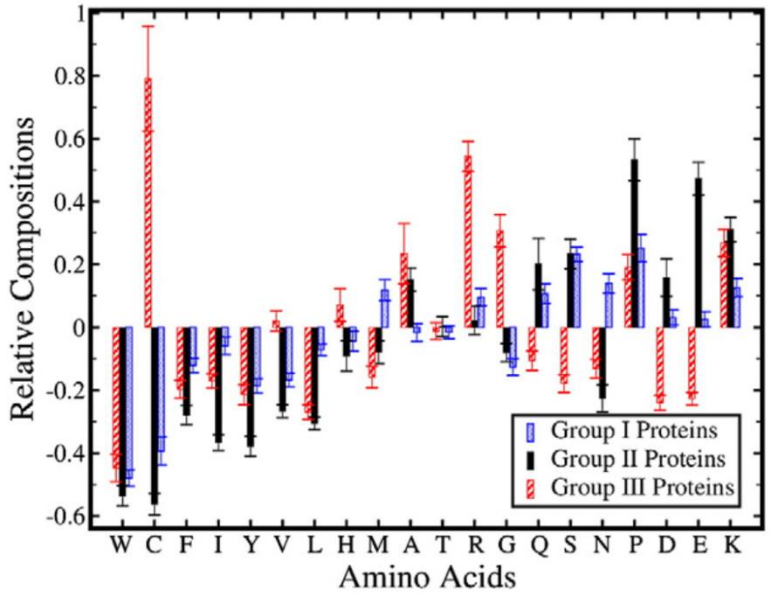


Figure 1-3 – Amino acid composition bias in IDPs. The amino acid composition of disordered proteins in comparison to ordered proteins, reproduced from Rani, Baruah and Biswas (2014). Group 1 consists of 9508 disordered proteins taken from the PDB for which the crystal structure contained at least 3 consecutive residues which were missing electron density i.e. were unresolved. Group 2 consists of 91 disordered proteins taken from DisProt (Release 6.02) for which there was total disorder, with no residues resolved in either crystal structures or NMR. Group 3 consists of 83 disordered proteins taken from the PDB for which there was less than 5% total secondary structure but contained at least 40 well defined residues in the crystal structure.

1.2.1 Biophysical Properties of IDPs

IDPs are defined by their lack of stable structure. It has been known for many decades that a protein's structure is determined entirely by its sequence (Anfinsen, 1973), so therefore it follows that disorder must also be encoded at the sequence level. An early analysis of 275 natively folded and 91 natively unfolded proteins found that IDPs were enriched in Ala, Arg, Gln, Glu, Gly, Lys, Ser and Pro (so called "disorder-promoting" residues) and depleted Asn, Cys, Ile, Leu, Phe, Trp, Tyr and

Val (so called "order-promoting" residues) with Asp, His, Met and Thr being ambiguous (Dunker *et al.*, 2001). A recent, more exhaustive study of 1917 globular proteins and 9682 disordered proteins (disordered being defined more loosely, see paper) yielded similar results (see Figure 1-3) (Rani, Baruah and Biswas, 2014). Amino acid composition is by no means the only method of discovering intrinsic disorder in proteins and algorithms used in there detection use more than 50 different predictors of disorder (He *et al.*, 2009), including net charge, 14 Å contact number, total Arg+Glu+Ser+Pro content, β-sheet propensity, total Cys+Phe+Tyr+Trp content, and flexibility.

While it is broadly true that bulky, hydrophobic residues are disfavoured in IDPs and IDRs, they are very rarely excluded from IDPs and IDRs. Molecular recognition features (MoRFs) are typically enriched with hydrophobic and aromatic residues, particularly Phe and Tyr (Mohan *et al.*, 2006). Moreover aromatic residues in MoRFs typically are directly involved in the binding interaction (Vacic *et al.*, 2007). Consequently aromatic residues are sparse, but strategically placed in IDPs and IDRs to enable proper function (Fuxreiter, Tompa and Simon, 2007) (see Section 1.2.3).

Lacking any fixed structure, IDPs and IDRs instead exist as diffuse ensembles of structures, with atom and backbone Ramachandran angles varying over time without a specific equilibrium value. While these "protein clouds" (Dunker and Uversky, 2010) are highly dynamic they are often well described by a fairly limited number of lower-energy conformations (Choy and Forman-Kay, 2001; Huang and Stultz, 2008). Consequently IDPs can form both extended regions with diffuse, poorly packed side chains (random coil-like, see Figure 1-4a) and partially collapsed, compacted regions (pre-molten globule-like, see Figure 1-4b) with noticeable secondary structure (molten globule-like) or a mixture (see Figure 1-4c) (Uversky and Dunker, 2010).

Charge also plays a number of important roles in IDPs and IDRs as they tend to be enriched in charged residues. It is well understood from studies of polyglycine that the polypeptide backbone has an intrinsic preference for forming collapsed structures (Tran, Mao and Pappu, 2008) and this driving force has to be overcome in order for extended conformations to occur. The high density of charged residues (Uversky, Gillespie and Fink, 2000) is the counter driving force that can, by modulating the effects of both hydration and solute ion coordination (Bokor *et al.*, 2005; Csizmók *et al.*, 2005; Tompa *et al.*, 2006), determine whether IDPs form molten globule-like, pre-molten globule-like or random coil-like structures (see Figure 1-4). Additionally Das and Pappu (2013) have shown that not only the number, but the distribution of charged residues has a significant impact on the structures adopted by disordered proteins and have developed a software tool, CIDER, to predict the effect that charge distribution has on the conformations adopted by IDPs (Holehouse *et al.*, 2017). Charge also has significant effect on IDP function, discussed in Sections 1.2.2 and 1.2.3.

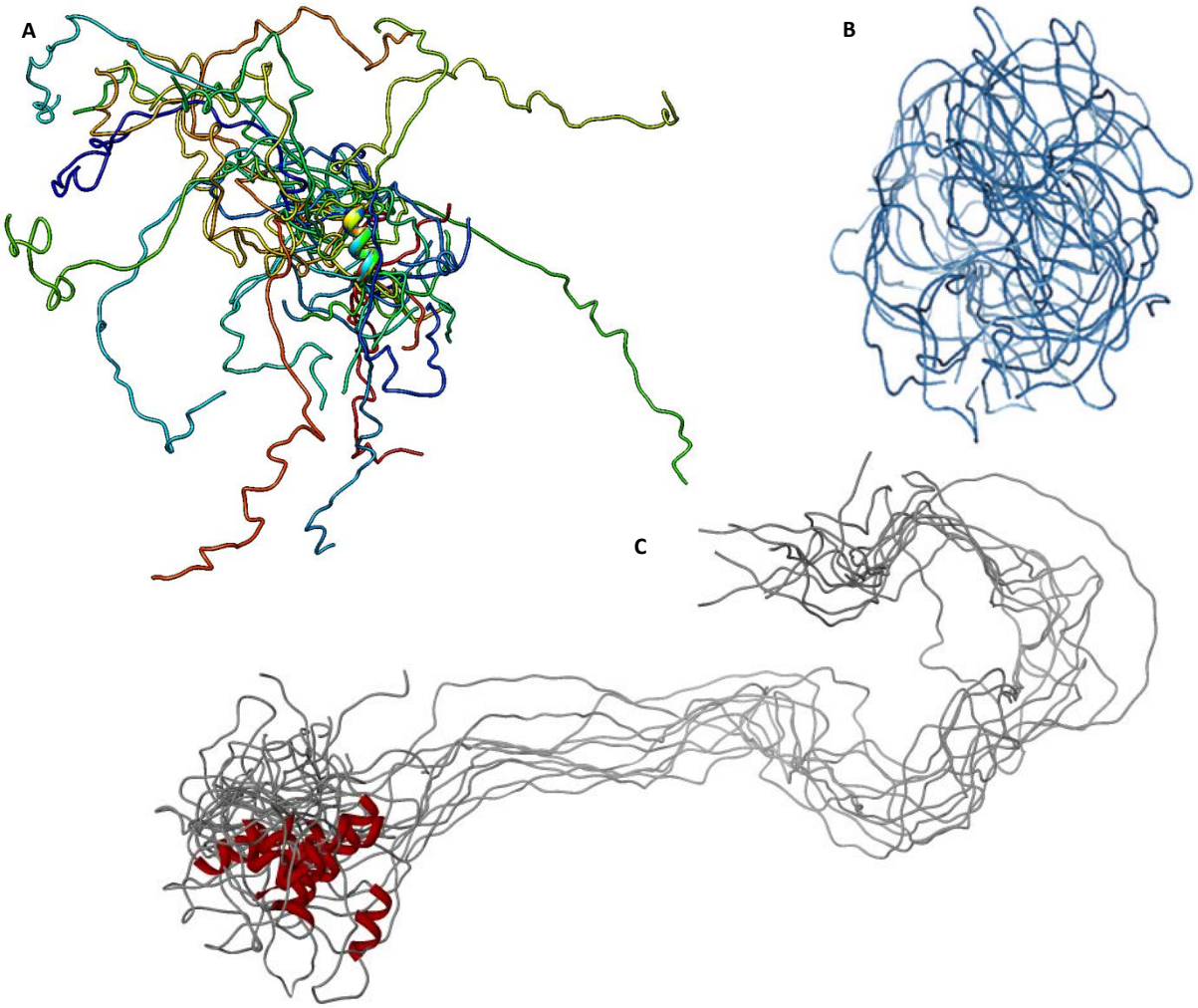


Figure 1-4 – IDP ensemble structures. Three examples of the ensembles of structures that IDPs adopt in solution. (A) is the ensemble (20 chains) adopted by thylakoid soluble phosphoprotein of 9 kDa (TSP9) as determined by solution NMR from PDB ID 2FFT (Song *et al.*, 2006) and rendered using UCSF Chimera. (B) is the ensemble adopted by activator of thyroid and retinoid receptors (ACTR) reproduced from Keppel *et al.* (2011). (C) is the ensemble for amelogenin as generated by MODELLER reproduced from Kozlowski and Bujnicki, (2012).

Many IDPs fold upon binding to their specific protein partners (Wright and Dyson, 2009; Keppel, Howard and Weis, 2011), even adopting different conformation depending on which partner they are binding to (Karush, 1950; Kriwacki *et al.*, 1996; Dunker *et al.*, 2001). For example the c-terminal regulation domain of tumour suppresser p53, which is disordered in its free form, can form a short α-helix, a β-strand or two differently shaped coils, depending on whether it is bound to s100B(ββ) (PDB ID 1DT7, Rustandi, Baldisseri and Weber, 2000), sirtuin (PDB ID 1MA3, Avalos *et al.*, 2002), cyclin A (PDB ID 1H26, Lowe *et al.*, 2002) or the CBP bromo domain (PDB ID 1JSP, Mujtaba *et al.*, 2004) respectively.

The flexibility of IDP also impacts its binding dynamics, their interactions with other protein partners tend to be high specificity/low affinity (Dunker *et al.*, 2001) with affinities typically falling in the low micromolar to low nanomolar range (Dogan, Gianni and Jemth, 2014). This low affinity/high specificity binding arises through the higher entropic cost of the disorder-to-order transitions IDPs

undergo on binding, compared to ordered proteins, which a recent computational study estimated at 1.0-1.5 kcal mol^{-1} per residue (Baxa *et al.*, 2014). Yet IDP-folded protein complexes are typically only 2.5 kcal mol^{-1} weaker than folded-folded complexes (Teilum, Olsen and Kragelund, 2015), suggesting that the interactions of IDPs with folded partners is actually more efficient. This is not to say that IDPs are incapable of high affinity binding, affinities in the high to low picomolar range have been observed (S-peptide-S-protein, $K_d \approx 6$ pM (Goldberg and Baldwin, 1998), PUMA BH3-Mcl1 , $K_d \approx$ 100 pM (Rogers, Steward and Clarke, 2013)). This usually arises through "fuzzy" binding, where disorder is partially maintained and the entropy cost minimised, discussed in Section 1.2.2.

Many factors within the cellular environment can affect IDP conformation. Factors such as salt concentration and pH have a significant impact on IDP conformation, particularly as a consequence of their high proportion of charged residues. Another factor that is important when comparing *in vitro* and *in vivo* studies is the effect of molecular crowding which can affect IDPs dramatically, many gaining significant partial structure in crowded environments. Some IDPs, such as α-synuclein (McNulty, Young and Pielak, 2006), remain completely unfolded in the crowded cellular environment (Szasz *et al.*, 2011), whereas other proteins, such as bacterial protein FlgM (Dedmon *et al.*, 2002), appear disordered *in vitro*, but are fully folded in their native cellular environments.

As has already been noted (see Section 1.1.2), the ability of tandem repeat proteins to dynamically partially unfold provides an evolutionary advantage by facilitating regulation through PTMs. IDPs are no less amenable to PTMs where they can have a significant impact not only on their interactions with other proteins and ligands but can induce dramatic conformational changes (Bah and Forman-Kay, 2016), leading complete folding (Bah *et al.*, 2014), oligomerisation (Mitrea *et al.*, 2014; Levine *et al.*, 2015) and phase-separation (P. Li *et al.*, 2012).

1.2.2 Binding Modes of IDPs

The inherent properties of IDPs and IDRs have some interesting effects on their ability to interact with other proteins and ligands as a given disordered region can contain any number of different MoRFs (Mohan *et al.*, 2006; Vacic *et al.*, 2007). This can either be in the form linear motifs in the protein sequence e.g. SLIMs (reviewed by Davey, Cyert and Moses (2015)), or secondary structural elements that are conformationally selected for or which fold upon binding (Wright and Dyson, 2009). Consequently IDP binding can be promiscuous, where a single IDP can bind to more than one partner, such as p53 (described above) or the "Janus" Chaperones (Tompa and Kovacs, 2010), or several IDPs can bind to a single ordered (or disordered) partner e.g. β-catenin (tandem repeat protein) binding to TCF/LEF family transcription factors, E-cadherin, axin and adenomatous polyposis coli (APC) (all IDPs, described in more detail in Section 1.5).

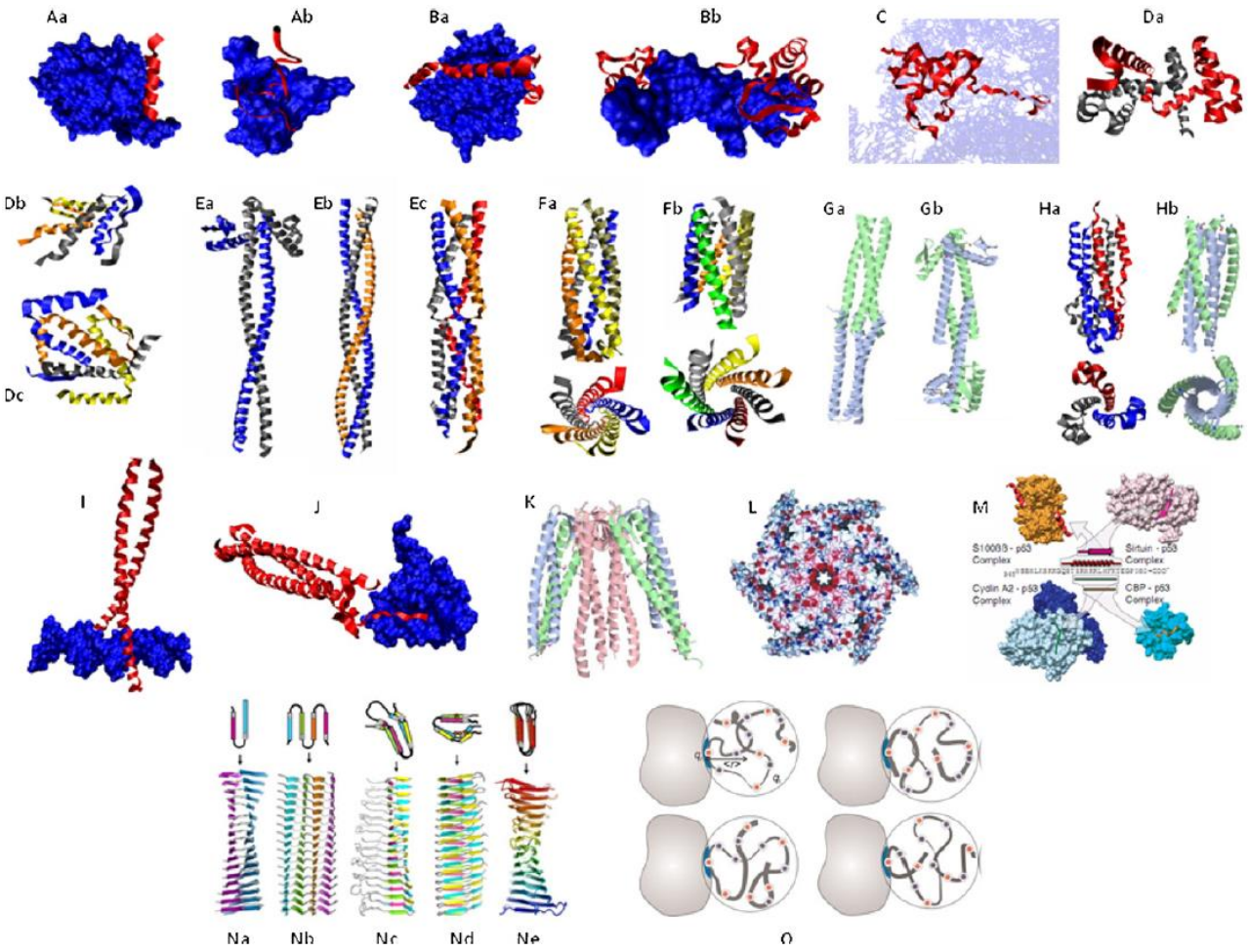


Figure 1-5 – A portrait gallery of IDPs in complex. Reproduced from Uversky (2011a). Description taken from that paper: *"**A. MoRFs**. **a**, α-MoRF, a complex between the botulinum neurotoxin (red helix) and its receptor (a blue cloud) (PDB ID: 2NM1); **b**, i-MoRF, a complex between an 18-mer cognate peptide derived from the α1 subunit of the nicotinic acetylcholine receptor from* Torpedo californica *(red helix) and α-cobratoxin (a blue cloud) (PDB ID: 1LXH). **B. Wrappers**. **a**, rat PP1 (blue cloud) complexed with mouse inhibitor-2 (red helices) (PDB ID: 2O8A); **b**, a complex between the paired domain from the Drosophila paired (prd) protein and DNA (PDB ID: 1PDN). **C. Penetrator**. Ribosomal protein s12 embedded into the rRNA (PDB ID: 1N34). **D. Huggers**. **a**, E. coli trp repressor dimer (PDB ID: 1ZT9); **b**, tetramerization domain of p53 (PDB ID: 1PES); **c**, tetramerization domain of p73 (PDB ID: 2WQI). **E. Intertwined strings**. **a**, dimeric coiled-coil, a basic coiled-coil protein from* Eubacterium eligens *ATCC 27750 (PDB ID: 3HNW); **b**, trimeric coiled-coil, salmonella trimeric autotransporter adhesin, SadA (PDB ID: 2WPQ); **c**, tetrameric coiled-coil, the virion-associated protein P3 from Caulimovirus (PDB ID: 2O1J). **F. Long cylindrical containers**. **a**, pentameric coiled coil, side and top views of the assembly domain of cartilage oligomeric matrix protein (PDB ID: 1FBM); **b**, side and top views of the seven-helix coiled-coil, engineered version of the GCN4 leucine zipper (PDB ID: 2HY6). **G. Connectors**. **a**, human heat shock factor binding protein 1 (PDB ID: 3CI9); **b**, the bacterial cell division protein ZapA from* Pseudomonas aeruginosa *(PDB ID: 1W2E). **H. Armature**. **a**, side and top views of the envelope glycoprotein GP2 from Ebola virus (PDB ID: 2EBO); **b**, side and top views of a complex between the N- and C-terminal peptides derived from the membrane fusion protein of the Visna (PDB ID: 1JEK). **I. Tweezers or forceps**. A complex between c-Jun, c-Fos and DNA. Proteins are shown as red helices, whereas DNA is shown as a blue cloud (PDB ID: 1FOS). **J. Grabbers**. Structure of the complex between βPIX coiled coil (red helices) and Shank PDZ (blue cloud) (PDB ID: 3L4F). **K. Tentacles**. Structure of the hexameric molecular chaperone prefoldin from the archaeum* Methanobacterium thermoautotrophicum *(PDB ID: 1FXK). **L. Pullers**. Structure of the ClpB chaperone from* Thermus thermophilus *(PDB ID: 1QVR). **M. Chameleons**. The C-terminal fragment of p53 gains different types of secondary structure in complexes with four different binding partners, cyclinA (PDB ID: 1H26), sirtuin (PDB ID: 1MA3), CBP bromo domain (PDB ID: 1JSP), and s100B(ββ) (PDB ID: 1DT7). **N. Stackers or β-arcs**. **a**, stack of β-arches, β-amyloid; **b**, superpleated β-structure (Sup35p, Ure2P, α-synuclein); **c**, stack of β-solenoids (prion); **d**, stack of β-arch dimers (insulin); **e**, β-solenoids. Modified from Kajava, Baxa and Steven (2010). **O. Dynamic complexes**. Schematic representation of the polyelectrostatic model of Sic1-Cdc4 interaction (Mittag et al., 2008). Schematic of an IDP (ribbon) interacting with a folded receptor (gray shape) through several distinct binding motifs and an ensemble of conformations (indicated by four representations of the interaction). The intrinsically disordered protein possesses positive and negative charges (depicted as blue and red circles, respectively) giving rise to a net charge q_i, while the binding site in the receptor (light blue) has a charge q_r. The effective distance <r> is between the binding site and the centre of mass of the intrinsically disordered protein. Reproduced from Mittag, Kay and Forman-Kay (2009). This figure is based on the data represented in Uversky (2011b)."*

Due to their unique capacity to adjust their structures to any given binding partner, many of which are unattainable to ordered proteins, and ability to present virtually any MoRF, the vast array of structures IDPs have been observed adopting in complex should be unsurprising. Broadly, these can be divided into static complexes (Uversky, 2011b), where the IDP is ordered in complex (and can often be seen in X-ray crystal structures), and dynamic or "fuzzy" complexes, where the IDP retains a significant proportion (sometimes all) of its disorder whilst in complex (Borg *et al.*, 2007; Sigalov, Zhuravleva and Orekhov, 2007; Mittag *et al.*, 2008; Mittag, Kay and Forman-Kay, 2009). Uversky (2011b) has subdivided the static complexes into 14 subgroups: MoRFs, wrappers, chameleons, penetrators, huggers, intertwined strings, long cylindrical containers, connectors, armature, tweezers and forceps, grabbers, tentacles, pullers, and stackers or β-arcs (see Figure 1-5 and the paper).

Mechanistically, IDP binding is equally varied, and recently Olsen, Teilum and Kragelund (2017) grouped them into four distinct modes of binding: Simple two-state binding, Avidity, Allovalency and Fuzzy binding. Simple two-state binding is, as the name implies, the simple case where binding can be approximated to single MoRF binding to a single receptor site on the binding partner (see Figure 1-6a). This is the most frequent case for conformational selection (Wright and Dyson, 2009), where the IDP samples a range of structures but only binds to its partner when the right conformation is adopted. It can also be the case for folding on binding interactions, if the whole process is highly concerted, as is the case for PUMA binding to MCL-1 (Rogers, Wong and Clarke, 2014).

Avidity arises when there are two or more independent MoRFs on the ligand IDP and an equal number of unique, complimentary receptor sites on the binding partner, which cannot exchange (see Figure 1-6b). In this model, the individual binding epitopes can bind separately to their cognate receptor sites and are separated by a disordered linker that ensures that once one site is bound the other epitopes are held close in space to their corresponding receptor sites. This behaviour greatly increases the probability of the subsequent binding events occurring after the first, lowering the entropic cost of binding more than one ligand, introducing cooperativity (Kitov and Bundle, 2003). However this disordered linker needs to be reasonably short otherwise the adjacent binding sites become decoupled and the effect is lost (Bobrovnik, 2007). Avidity was originally a term used to describe antibody-antigen binding, which Olsen, Teilum and Kragelund (2017) subsequently co-opted to describe this particular mode of IDP binding (Crothers and Metzger, 1972).

Allovalency is a concept originally developed by Klein, Pawson and Tyers (2003) and later expanded upon by Levchenko (2003) and Locasale (2008). Allovalency differs from avidity, as for avidity there are n unique matched epitope-receptor pairs, whereas as in allovalency, n (nearly) identical binding

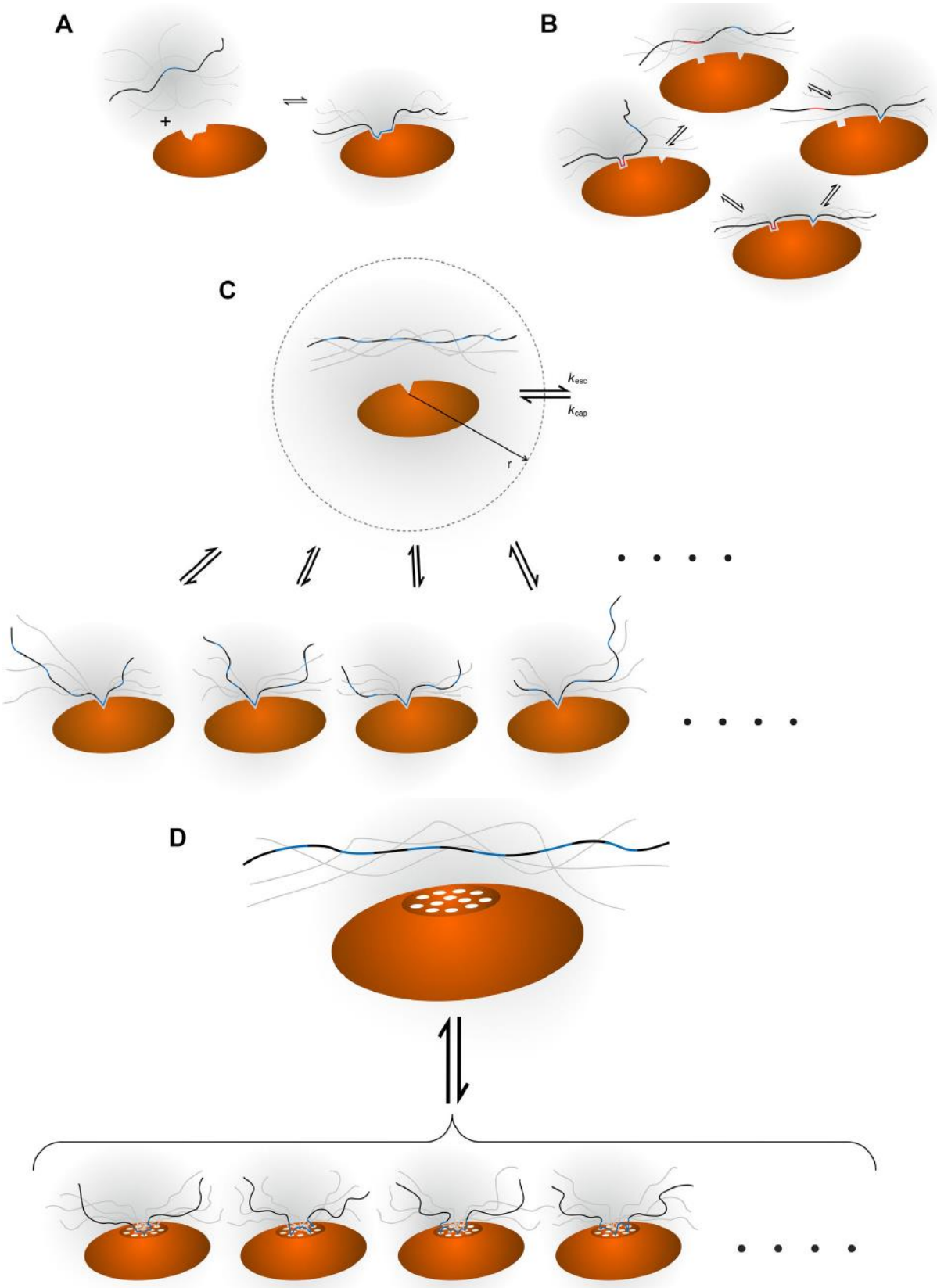


Figure 1-6 – An illustration of the four different binding modes of IDPs. (A), (B), (C) and (D) are cartoon diagrams of simple two-state, avidity, allovalency and fuzzy binding respectively. The IDP ligand is shown in grey with its binding epitopes highlighted in blue and red, and its macromolecular receptor is shown in orange. For (C) the dashed grey circle represents the capture radius of the receptor and for (D) the white dots indicate the cluster of micro -receptors sites with which the IDP interacts. Reproduced from Olsen, Teilum and Kragelund (2017).

epitopes compete for a single receptor site, which can only be occupied by a single epitope at a time

(see Figure 1-6c). The individual epitopes may have very low individual affinities for the binding site, but the presence of multiple epitopes held with a "capture radius" increases the overall affinity (Mittag *et al.*, 2008). Unlike with simple two-state binding and avidity, allovalency often does not result in long-lived static complexes, combined with the array of adoptable structures inhibiting crystal packing, it is difficult to obtain crystal structures for allovalent complexes unless smaller fragments containing only a few epitopes are used (Xing *et al.*, 2004)

The definitive example of an allovalent complex is the complex formed by the cyclin-dependent kinase inhibitor Sic1 and its binding partner, the F-box protein Cdc4 (Klein, Pawson and Tyers, 2003; Borg *et al.*, 2007; Mittag *et al.*, 2008). Sic1 contains nine suboptimal binding epitopes for a single binding site on Cdc4, each of which can be "activated" by phosphorylation of either a serine or threonine residue within the epitope (Klein, Pawson and Tyers, 2003). In order to achieve its high affinity binding, at least six of the nine epitopes must be phosphorylated but which epitopes is entirely arbitrary (Nash *et al.*, 2001; Borg *et al.*, 2007; Mittag *et al.*, 2008).

The final binding mechanism proposed by Olsen *et al*. is fuzzy binding. Inspired by the mathematical term "fuzzy logic", whereby the true answer to a question can be no (0), yes (1) or any value in between, the concept was first introduced by Tompa and Fuxreiter in 2008. Fuxreiter *et al.* would later describe fuzzy complexes as *'protein complexes, where conformation heterogeneity of IDRs is retained and is required for function'* (Sharma *et al.*, 2015), highlighting the key distinction between a fuzzy complex and non-specific interactions: that fuzzy complexes have a biological consequence. Mechanistically, fuzzy complexes are best described as an IDP-receptor complex comprised of a large number of smaller functional groups forming many transient interactions with a large number of micro-receptor sites. These functional groups can be any normally seen in proteins e.g. –OH, $–CH_3$, $–CO_2^-$, $–NH_3^+$, –aromatic ring, etc., but the key is that these small interactions are individually weak and bind and unbind very rapidly, allowing the IDP to continue sampling a relatively large conformational space and reduce the entropy cost of binding. The earliest known example of fuzziness is the T cell receptor ζ-subunit recognising SIV pathogenicity factor Nef (Sigalov *et al.*, 2008). Fuzzy complexes in general cannot be captured in X-ray crystal structures and requires a combination of techniques such as SAXS, NMR, CD and hydrogen-deuterium exchange mass spectroscopy (HDX-MS) to obtain an accurate picture of the complex ensemble (Gógl *et al.*, 2016; Peti *et al.*, 2018). Moreover as only a single state is allowed in a crystal lattice, this can produce misleading artefacts.

Independent from their mode of binding, Tompa and Csermely grouped IDPs into six broad functional classes, based on their mode of action: Assemblers, chaperones, display sites, effectors,

entropic chains and scavengers (Tompa, 2002, 2003; Tompa and Csermely, 2004). Chaperones aid the folding/reroute the misfolding of other macromolecules, and are often capable of binding many different macromolecules in a range of intermediate folded states. Intrinsic disorder provides chaperones the flexibility to adapt to the structure of different partially or misfolded targets as well as aiding solubility, increasing the "first-contact" radius (sometimes referred to as "flycasting") and/or acting through entropic transfer (see Tompa and Csermely (2004)). Display sites act as recognition sites for other enzymes, usually for post-transcriptional modification and/or limited proteolysis. Effectors modulate the actions of other proteins, usually as an inhibitor or activator, typically by binding to them. Entropic chains are more loosely defined as any function that simply could not be performed by a protein with a well-defined compact structure such as spacer/linker regions, entropic gating and molecular springs. Scavengers bind, store and sometime neutralise small molecules. Where only a short and simple binding motif is required to bind the target small molecule, it is advantageous to have an extended, multivalent binding surface that can sample a large amount of space quickly to find and sequester the target. Finally assemblers use their intrinsic disorder to help coordinate and stabilise macromolecular complexes, either by acting as flexible linkers between structured binding domains, or in a similar mode to scavengers: presenting multiple binding motifs and sampling a large conformational space to locate the required complex components, as well as providing additional stabilisation to the holo-complex once assembled.

1.2.3 IDPs in Physiological and Pathological Functions

IDPs have been observed performing pretty much every role within cells (Xie *et al.*, 2007a, 2007b, 2007c; Sharma *et al.*, 2015), but due to their high-specificity/low-affinity binding and their promiscuous one-to-many and many-to-one binding capabilities they are more often seen acting as interaction hubs in protein networks (Dunker *et al.*, 2005; Uversky, Oldfield and Dunker, 2005; Haynes *et al.*, 2006; Singh, Ganapathi and Dash, 2006). This presents a problem as hubs are central to normal function and signalling pathway cross-talk in all living organisms and disruption of hubs is frequently lethal (Jeong *et al.*, 2001) or results in serious pathologies, particularly cancers and neurodegenerative diseases (Iakoucheva *et al.*, 2002; Tompa, 2012). Some better known examples of IDP hubs in humans include p53 (cancer, Meek (2015)), α-synuclein and tau protein (Parkinson's diasease and Alzheimer's disease respectively, Moussaud *et al.* (2014)), myelin basic protein (multiple sclerosis, Zhou *et al.* (2017)) and BRCA1 and estrogen receptor α (breast cancer, Gaboriau *et al.* (2015) and Jahandoost, Farhanghian and Abbasi, (2017) respectively).

As a consequence, the prevalence of intrinsic disorder in disease gave rise to the D^2 concept (disorder in disorders), where it is suggested that disorder is prevalent in disorders precisely due to

the unusual properties of IDPs (Uversky, Oldfield and Dunker, 2008). In globular proteins, mutation that destabilise their folded structure results in their unfolding, leading to degradation before a mutant signal can be produced. However IDPs are biologically active in their disordered form (Hammoudeh *et al.*, 2009), so mutation within IDRs are not immediately recognised by the proteasome (if at all), resulting in prolonged aberrant signalling, leading to disease. Additionally, as IDPs play key roles in regulation and signalling pathways, altered availability (either by over or underexpression) can also have pathological results (Babu *et al.*, 2011).

Table 1-3: Normal and pathological functions of IDPs in humans

IDP or IDR Containing Protein	Normal Function	Associated Pathology in Humans
p53	Tumour suppressor - activates DNA repair proteins, cell cycle arrest (G1/S checkpoint), angiogenesis inhibitor, initiator of apoptosis	Tumour cell proliferation
α-synuclein	Synaptic vesicle trafficking (not well understood)	Alzheimer's Disease (via amyloidogenesis)
HGMA (high mobility group A)	Tumour suppressor - cell growth, proliferation, differentiation, and apoptosis, particularly embryo development	Tumour initiation, cell progression, metastasis and invasiveness
estrogen receptor α	Estrogen sex hormone signalling	Breast cancer in men and women; Gynecomastia in men; Estrogen Insensitivity Syndrome - Tall stature, osteoporosis and infertility in both men and women, abnormal development in puberty and many other problems in women
Arf and Mdm2	Regulators of p53 (see p53)	Delayed DNA break repair, chromosomal abnormality and genome instability
measles virus nucleoprotein (MVN)	In the Measles virus: nucleocaspid assembly and ssRNA genome encapsidation	Measles
Axin and adenomatous polyposis coli	Structuring the β-catenin destruction complex (see Section 1.5.2)	Aberrant Wnt signalling leading to colorectal cancers (see Section 1.5.2)
BRCA1	Tumour suppressor - DNA damage signalling and repair, centrosome duplication, growth regulation and apoptosis	Breast cancer
Caldesmon and calmodulin	Regulation of smooth muscle contraction	Loss of smooth muscle tone, essential for normal heart contractility

Table 1-3 (cont.)

IDP or IDR Containing Protein	Normal Function	Associated Pathology in Humans
p27 and $p21^{Waf1/Cip1/Sdi1}$	Cyclin-dependent kinase inhibitor	Tumour cell proliferation
XPA	Scaffold for DNA nucleotide excision repair proteins	Xerderma pigmentosum (hypersensitivity to UV, increasing skin cancer risk)
tau protein	Actin microtubule stabilisation	Multiple tauopathies resulting in neurodegeneration, notably Alzheimer's disease
Amyloid β	As amyloid precursor protein: synapse formation, neural plasticity and iron export	As amyloid β: Alzheimer's Disease
Prion Proteins	Unclear. Evidence for maintenance of long-term memory and stem cell renewal	Assorted transmissible spongiform encephalopathies including Creutzfeldt-Jakob disease (human "mad cow disease"), Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia and Kuru
IκBα	Inhibits nuclear translocation of NF-κB transcription factors	Hodgkin's lymphoma
myelin basic protein	Formation of myelin sheaths in the nervous system	Multiple sclerosis
Strathmin	Cytoskeletal regulation via inhibition of tubulin polymerisation	Cancer - uncontrolled cell division through inability to prevent mitotic spindle assembly
Huntingtin	Unclear. Absence is embryonic lethal	Huntington's disease
PTEN	Phosphorylase for PIP3	Numerous cancers, particularly prostate cancer in men

Proteins listed have been pooled from papers cited in Section 1.2

What is more interesting is the range of functions IDPs are not typically observed performing. By analysing the Swiss-Prot database Xie *et al.* found that only in five major functional categories where IDPs were significantly under represented (Xie *et al.*, 2007a), those being catalysis, transport, biosynthesis, metabolism and trans-membrane proteins. Even so it is possible to find examples of proteins performing these functions where intrinsic disorder is still essential to function.

Retinaldehyde Dehydrogenase II (RalDH2) in a member of the NAD-dependant aldehyde dehydrogenase (ALDH) superfamily, which relies on intrinsic disorder for its function (Bordelon *et al.*, 2004). Unlike other members of this family (Liu *et al.*, 1997; Steinmetz *et al.*, 1997; Johansson *et al.*, 1998; Moore *et al.*, 1998), the crystal structures of RalDH2 reveal that one wall of the catalytic

pocket is disordered (Lamb and Newcomer, 1999) and undergoes a disorder-to-order transition on substrate binding. Bordelon *et al*. (2004) were able to show through mutagenesis that this disordered region has an effect on catalysis beyond simple substrate binding affinity. They rationalise that because the primary substrate for RalDH2, retinal, is much larger the primary substrates of other members of the family, having a disordered wall instead of a rigidly structured one allows a gating mechanism. Rather than maintaining a large open pocket, the disordered region acts to exclude water and other aldehydes from the catalytic site as well adjusting to the optimal catalytic conformation once the correct substrate is bound.

IDPs involved in transport often act as scavengers, where multiple short, simple binding motifs are arranged in tandem in an extended disordered region, acting as a multivalent binding surface that can sample a large amount of space quickly to find and sequester the target molecules (Tompa, 2002, 2003). For example the 69-residue mitochondrial protein Cox17 plays a vital role in importing copper (I) ions into the mitochondria for the cytochrome-c oxidase complex (Terziyska *et al.*, 2005a, 2005b; Cobine, Pierrel and Winge, 2006). In the reducing environment of the cytoplasm Cox17 is disordered and capable of binding four copper (I) ions cooperatively (Abajian *et al.*, 2004), however when Cox17 is imported into the highly oxidising intermembrane space it folds via the formation of disulfide bonds (Banci, Bertini, Ciofi-Baffoni, Janicka, *et al.*, 2008; Banci *et al.*, 2011). In its oxidised folded state Cox17 is only capable of binding one copper (I) ion transferring the other three to Cox11 and Soc1 (Horng *et al.*, 2004; Banci, Bertini, Ciofi-Baffoni, Hadjiloi, *et al.*, 2008).

IDPs and IDRs involved in biosynthetic and metabolic processes tend not to be involved directly in the synthetic machinery; instead they tend to be involved in complex assembly. Examples include the L7 and L12 ribosomal proteins which play a vital role in coordinating the translation factors; EF-G, IF2 and RF3 in the bacteria 70S ribosome (Carlson *et al.*, 2017). They also can act as modulators of biosynthetic machinery such as CP12 in chloroplasts, which acts as a light sensitive switch for the Calvin cycle (the carbon-capture cycle that assimilates CO_2) (minireviewed by Gontero and Maberly (2012)). In the absence of light, the chloroplast environment is oxidising and CP12 forms a complex with the enzymes glyceraldehyde-3-phosphate dehydrogenase and phosphoribulokinase, inhibiting both (Wedel, Soll and Paap, 1997; Wedel and Soll, 1998). In the presence of light, ATP and NADPH are produced in the primary phase of photosynthesis, which turns the chloroplast into a reducing environment, CP12 becomes reduced, releasing the two enzymes to perform their function (Thieulin-Pardo *et al.*, 2015).

Transmembrane proteins also tend to require structure for function, as they are anchored within a membrane. However one particularly well studied example of IDPs in transmembrane proteins are the phenylalanine-glycine repeat-containing nucleoporins (FG Nups) which line the inner surface of nuclear pore complexes (NPCs) (reviewed by Sakiyama, Panatala and Lim (2017)). The FG Nups form a bristly, sieve-like hydrogel that prevents the translocation of large non-specific macromolecules above ≈ 40 kDa while still being permeable to smaller molecules (Rout *et al.*, 2003; Yamada *et al.*, 2010; Popken *et al.*, 2015). Selective transport across the nuclear envelope is facilitated by the importin family of proteins, sometimes called karyopherins (which are ARM repeat proteins) (Chook and Süel, 2011). Importin β1 binds to cargo proteins containing a nuclear localisation signal and the Importin-cargo complex translocates through the NPC in what appears to be a diffusion process (Yang, Gelles and Musser, 2004) but is actually directionally regulated by Ran GTPase (Görlich *et al.*, 1996). It is still unclear precisely how this fast, selective, transport works mechanistically but a possible explanation is that importin β1 has approximately ten hydrophobic grooves which could potentially bind to FG repeats (Isgro and Schulten, 2005) in a manner somewhere between allovency and fuzzy binding (Schoch, Kapinos and Lim, 2012; Kapinos *et al.*, 2014; Hough *et al.*, 2015; Milles *et al.*, 2015), where the effective capture radius of the complex is the entire inner pore (Olsen, Teilum and Kragelund, 2017).

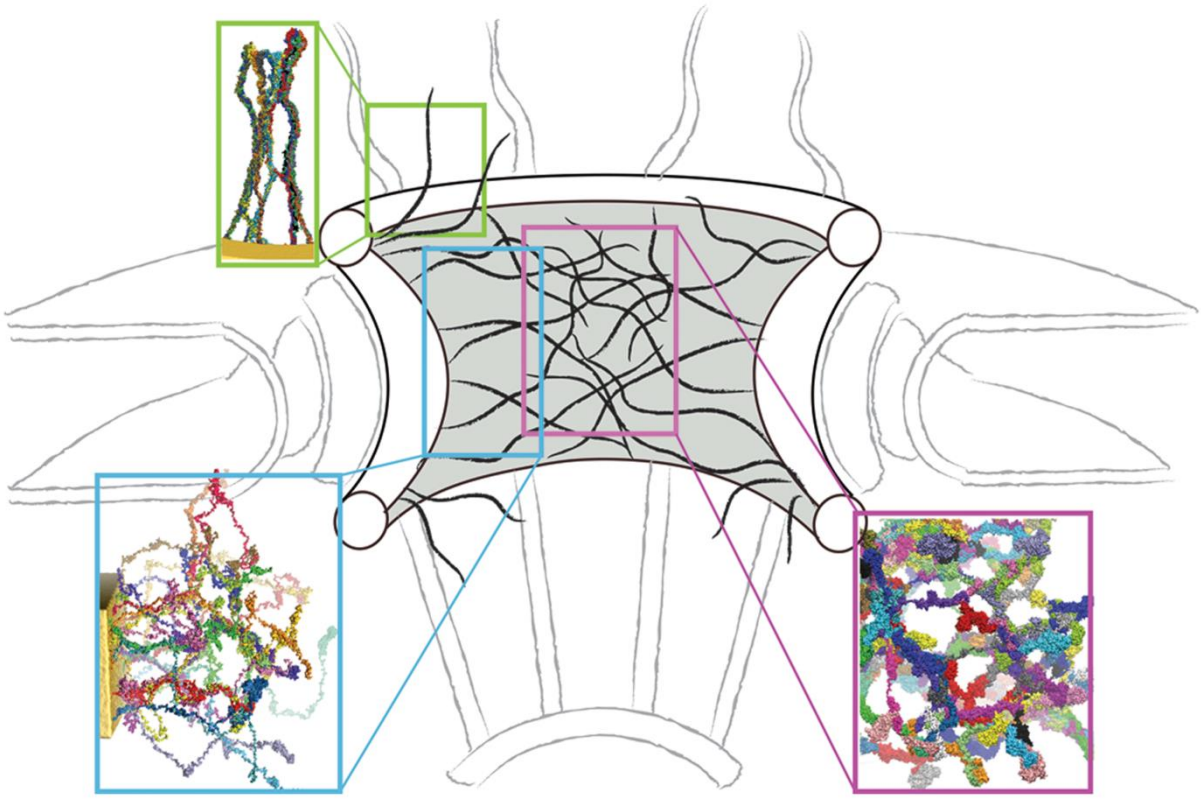


Figure 1-7 – Schematic model of a nuclear pore complex. The black strands represent the intrinsically disordered FG nucleoporin, Nsp1, within the pore of the nuclear pore complex. The green, blue and pink boxes contain stills taken from simulations of Nsp1 arranged in a ring array. Reproduced from Gamini *et al.* (2014).

Disregarding the "structure dictates function" paradigm, IDPs have the potential to function in ways that would be impossible for ordered proteins. One of the more interesting examples of unique functions of IDPs is the formation of membrane-less organelles, where soluble macromolecules undergo liquid-liquid phase separation (reviewed by Aumiller and Keating (2017)). The intrinsic properties of IDPs make them well-suited for the formation of membraneless organelles (reviewed by Uversky (2017a)). IDPs are conformationally flexible, capable of forming very elongated structures (Das and Pappu, 2013) and multivalent, allowing them to self-polymerise through disperse repeated units (e.g. alternating blocks of opposite charge (Nott *et al.*, 2015) or complementary donor-acceptor units, separated by flexible linkers (P. Li *et al.*, 2012)) as well as recruit other proteins that facilitate the organelles function. They are amenable to many forms of post-transcriptional modification (Wang *et al.*, 2014; Nott *et al.*, 2015), which may induce dynamic compartmentalisation in response to environmental changes and other stimuli (e.g. changes in salinity, temperature or binding partner concentration). Membraneless organelles constitute a deep and complex emerging field that is not directly relevant to the rest of this thesis, but I have mentioned them here as a fascinating example of the unique functionality of intrinsic disorder in proteins. For a more in depth discussion, as well as an extensive list of examples, I recommend the reviews by Aumiller and Keating (2017) and Uversky (2017a, 2017b).

1.3 The Canonical Wnt Signalling Pathway

The canonical, β-catenin-dependant Wnt signalling pathway is one of three signalling pathways activated by the secreted glyco-/lipoproteins of the Wnt family (Port and Basler, 2010). The other two non-canonical (or β-catenin-independent) Wnt signalling pathways are the Wnt/PCP (planar cell polarity) pathway, which regulates cytoskeleton dynamics, and the Wnt/Ca^{2+} pathway, which regulates intracellular Ca^{2+} levels, act through G-protein coupled mechanisms (Gordon and Nusse, 2006). These pathways are critical to the regulation of cell proliferation, differentiation and fate decisions in embryonic development and tissue homeostasis in adults (Logan and Nusse, 2004; Guder *et al.*, 2006; Clevers and Nusse, 2012), particularly the canonical pathway for which deletion and continuous activation are both embryonic lethal (Logan and Nusse, 2004). The β-catenin-independent pathways are not relevant to this Thesis and so will not be discussed further here, although excellent reviews have been written by Seifert and Mlodzik (2007) and Kohn and Moon (2005) respectively.

In the absence of a Wnt signal, cytosolic β-catenin is continuously synthesised then sequestered and targeted for proteasomal degradation by a multi-protein complex called the β-catenin destruction complex (BDC) (MacDonald, Tamai and He, 2009). The BDC is comprised of five main proteins: two

structural proteins, Axin and adenomatous polyposis coli (APC), two kinases, glycogen synthase kinase 3β (GSK3β) and casein kinase 1α (CK1α) and a phosphatase, protein phosphatase 2A (PP2A). Here β-catenin is hyperphosphorylated at the N-terminal disordered region by the combined action of GSK3β and CK1α (both are required, Kimelman and Xu (2006)), allowing it to be recognised and ubiquitinated by the β-TrCP E3 ubiquitin ligase (Wu *et al.*, 2003) and then subsequently degraded by the proteasome.

Both Axin and APC are large proteins (862 and 2843 residues, respectively), containing large IDRs (residues 211-780 and 800-2843, respectively), which are critical to their function and to the activity of the BDC (Minde *et al.*, 2011; Noutsou *et al.*, 2011). Together these two proteins form a scaffold onto which the other components and β-catenin can bind. Axin and APC can both self-associate, APC can homodimerise via its N-terminal coiled-coil domain (residues 1-50) (Joslyn *et al.*, 1993) and Axin can homo-oligomerise via its C-terminal DIX domain (residues 780-862) (Kishida *et al.*, 1999), which can also hetero-oligomerise with other DIX domain-containing proteins (Wong *et al.*, 2004). Additionally, Axin contains one APC-binding RGS domain (88-211) which can bind to one of three SAMP repeats in APC (Eklof Spink, Fridman and Weis, 2001). This multiplicity of binding allows APC and Axin to theoretically form an infinitely large network which could potentially be considered a membraneless organelle (Kunttas-Tatli, Roberts and McCartney, 2014) (see Figure 1.8).

The IDR of Axin contains binding sites for all the members of the BDC required for β-catenin phosphorylation within its IDR (GSK3β and CK1α and β-catenin itself, Xue *et al.* (2013)) yet it has been shown that APC is required for BDC function (McCartney and Näthke, 2008; Kunttas-Tatli, Roberts and McCartney, 2014). APC coordinates PP2A in the complex via an ARM repeat domain and also has eleven binding sites for β-catenin within its IDR (seven 20-mer motifs and four 15-mer motifs, Minde *et al.* (2011)), each of which has a lower affinity for β-catenin than Axin (Choi, Huber and Weis, 2006; Liu *et al.*, 2006). However these eleven repeats can be phosphorylated by GSK3β and CK1α, which increases their affinities for β-catenin above that of Axin (Ha *et al.*, 2004; Liu *et al.*, 2006), suggesting a mechanism of cycling β-catenin between Axin and APC through phosphorylation and dephosphorylation (by PP2A Ha *et al.* (2004)) of these repeats. It has been suggested that having this many binding repeats in a large IDR allows APC to operate as a scavenger for β-catenin (Minde *et al.*, 2011), as well as playing a role in recruiting the β-TrCP ubiquitin E3 ligase, post β-catenin phosphorylation (Yang *et al.*, 2006; V. S. W. Li *et al.*, 2012).]

The APC-β-catenin interaction provides an interesting example of an IDR binding to an ordered partner. The eleven β-catenin binding motifs, modulated by phosphorylation, is reminiscent of the Sic1-Cdc4 allovalent binding mentioned in Section 1.2.2 (Mittag *et al.*, 2008). However, the different

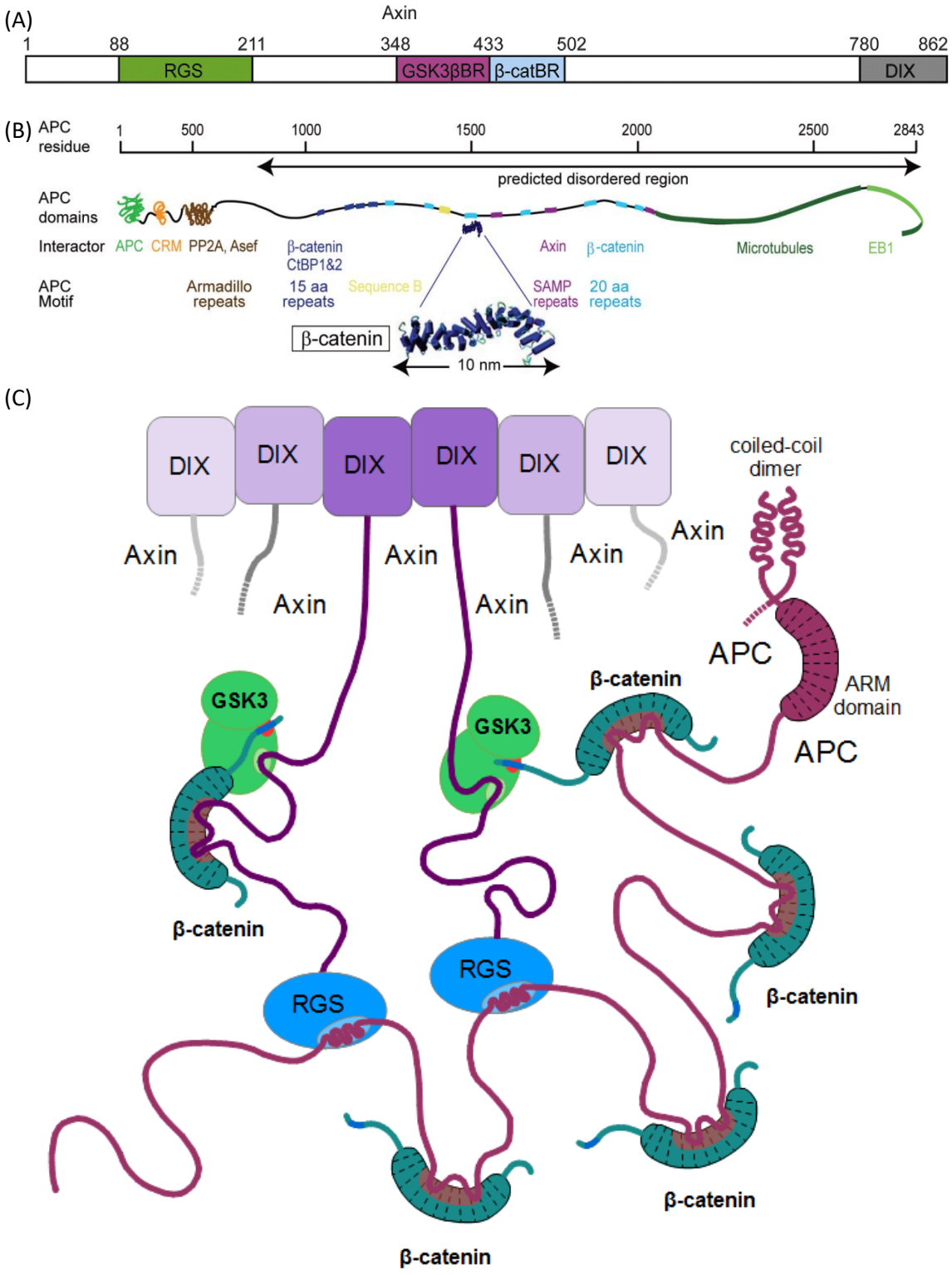


Figure 1-8 – Axin and APC in the β-catenin destruction complex. (A) and (B) show the positions of the domains of Axin and APC respectively, with the interaction partners for APC being highlighted in (B). (C) demonstrates the potential of Axin and APC to form an infinite network such as those seen in membraneless organelles. (A) and (B) were recreated from Noutsou *et al.* (2011) and Minde *et al.* (2011), respectively, and (C) was from Wikimedia commons.

sizes of the binding interfaces (8 residues for Sic1, >50 for APC) as well as the distance between repeats (<10 residues for Sic1, ≈100 for APC) suggests that APC-β-catenin binding does not follow the allovalency mechanism. Moreover currently unpublished work by Dr. P.J.E. Rowling in our research group suggests that β-catenin binding to an individual APC repeat follows a two-site avidity mechanism.

When the Wnt signal is given, the Wnt ligand binds to the transmembrane Frizzled (Fz) receptor and a co-receptor, LRP5 or 6 (usually 6), initiating a multistep process that leads to the BDC becoming plasma membrane-bound and sequestered, effectively inactivating it (Taelman *et al.*, 2010; Berwick and Harvey, 2012; Kim, Kim and Jho, 2013). This in turn leads to the accumulation of cytosolic β-catenin which is then able to enter the nucleus and bind to and co-activate the TCF/LEF family of transcription factors, activating the transcription of the Wnt target genes (see Figure 1-9) (MacDonald, Tamai and He, 2009).

The actual process of BDC inactivation and sequestration is not fully understood owing in large parts to the complexity and diversity of its regulation and crosstalk with other signalling pathways (reviewed by Kim, Kim and Jho (2013)). Three events are known to be important for this process: hyperphosphorylation of the intercellular region of LRP5/6 (Tamai *et al.*, 2004), Dishevelled (Dvl)-mediated clustering of Fz/LRP receptors at the plasma membrane (Zeng *et al.*, 2008), and endocytosis of the Fz/LRP/BDC complex in multi-vesicular bodies (Taelman *et al.*, 2010).

LRP5 & LRP6 contain five PPP(S/T)P repeats in their cytosolic tail which creates a docking site for Axin, when all five are phosphorylated (Tamai *et al.*, 2004). However a number of different studies have shown that many different kinases are capable of facilitating this step, thereby allowing modulation and crosstalk between multiple signalling pathways (Kim, Kim and Jho, 2013). These include membrane-bound CK1γ (Davidson *et al.*, 2005), G-protein-coupled receptor kinase (GRK) 5/6 (Chen *et al.*, 2009), the cyclin-Y-dependent kinase PFTK (Davidson and Niehrs, 2010), MAPKs (Davidson *et al.*, 2009), and cytosolic GSK3β (Zeng *et al.*, 2005). Additionally, it has been shown that Axin-bound GSK3β is also capable of phosphorylating the LRP5/6 cytosolic tail, providing a possible signal amplification method when combined with Fz/LRP clustering.

The Dishevelled (Dvl) phosphoprotein plays a central role in both the canonical and the non-canonical Wnt signalling pathways, being the protein that both relays the Wnt signal to the cytosolic BDC and facilitates its binding to FZ/LRP at the cell membrane (Schwarz-Romond, Metcalfe and Bienz, 2007). Like Axin, Dvl contains a DIX domain, which allows it to both homo-oligomerise, and hetero-oligomerise with other DIX domain containing proteins (Kishida *et al.*, 1999; Wong *et al.*,

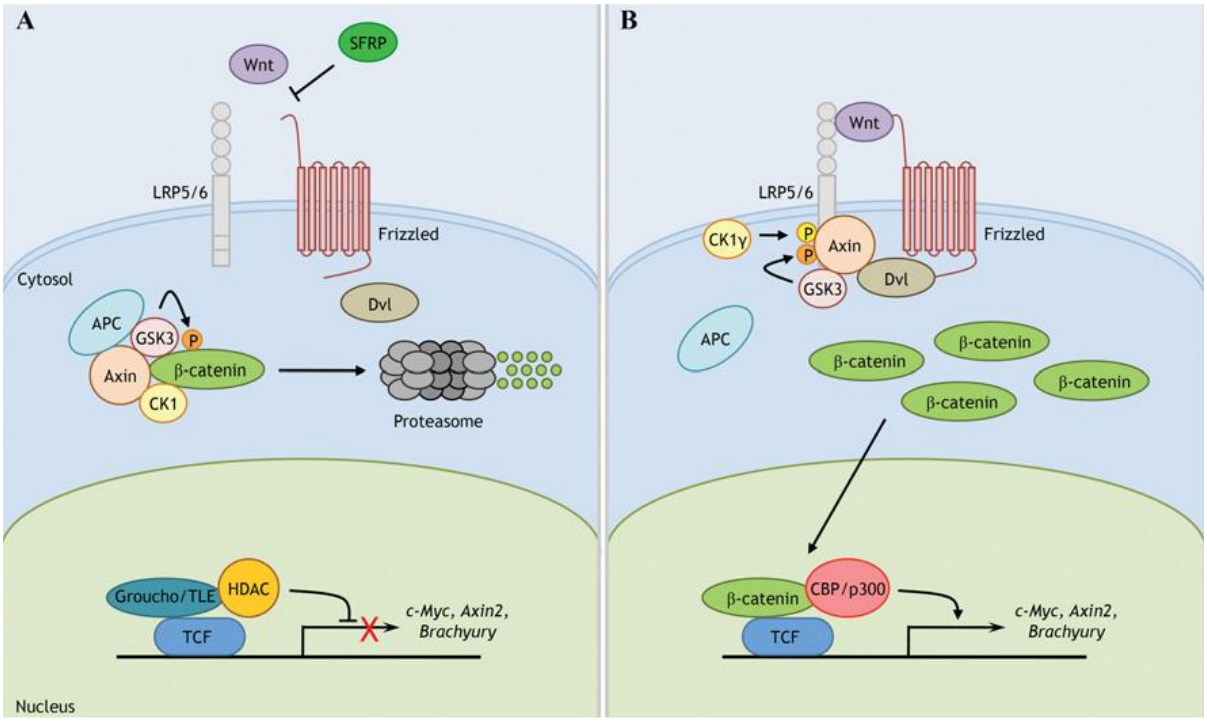


Figure 1-9 – The canonical Wnt signalling pathway. The canonical Wnt signalling pathway in the absence (A) and presence (B) of the Wnt ligand. In (A), β-catenin is continuously phosphorylated by the BDC (APC, Axin, CK1α and GSK3β), tagging it for degradation by the proteasome. Upon receiving the Wnt signal (B) the BDC is recruited to the cell membrane, freeing β-catenin, which translocates to the nucleus where it co-activates the TCF/LEF family of the transcription factors. Reproduced from Kim, Kim and Jho (2013).

2004), but Dvl also contains a PDZ domain, which can interact directly with the Fz receptor (Wong *et al.*, 2003; Punchihewa *et al.*, 2009). Precisely how Wnt stimulation induces a change in behaviour is complex and multifaceted but it is known to involve phosphorylation of Dvl by various kinases, including casein kinases, but not CK1α or γ (reviewed by Gao and Chen (2010) and Kim, Kim and Jho (2013)).

Here we can see the multiple pathways involved in BDC membrane recruitment as well as a mechanism for the clustering of receptors at the membrane. Axin (and the rest of the BDC) is recruited to the Fz/LRP receptor complex either to the phosphorylated LRP5/6 cytosolic tail (Tamai *et al.*, 2004) or to the intracellular domain of Fz using Dvl as an adaptor (Punchihewa *et al.*, 2009), leaving the other site free to bind a second Axin. Through this mechanism, one can imagine a "daisy-chaining" of receptors and Axin/Dvl hetero-oligomers, resulting in the clustering of receptors. Additionally, Dvl oligomers can bind multiple Fz/LRP receptor complexes with or without the presence of Axin providing another mode of receptor clustering.

Finally, the aggregated Wnt/Fz/Dvl/BDC complexes are endocytosed into multi-vesicular bodies, a step which has been shown to be absolutely essential to Wnt signalling (Blitzer and Nusse, 2006; Yamamoto, Komekado and Kikuchi, 2006; Bilic *et al.*, 2007) (see Figure 1-10). It is currently thought that this is a key step in inhibiting the activity of the BDC: physically excluding GSK3β from the

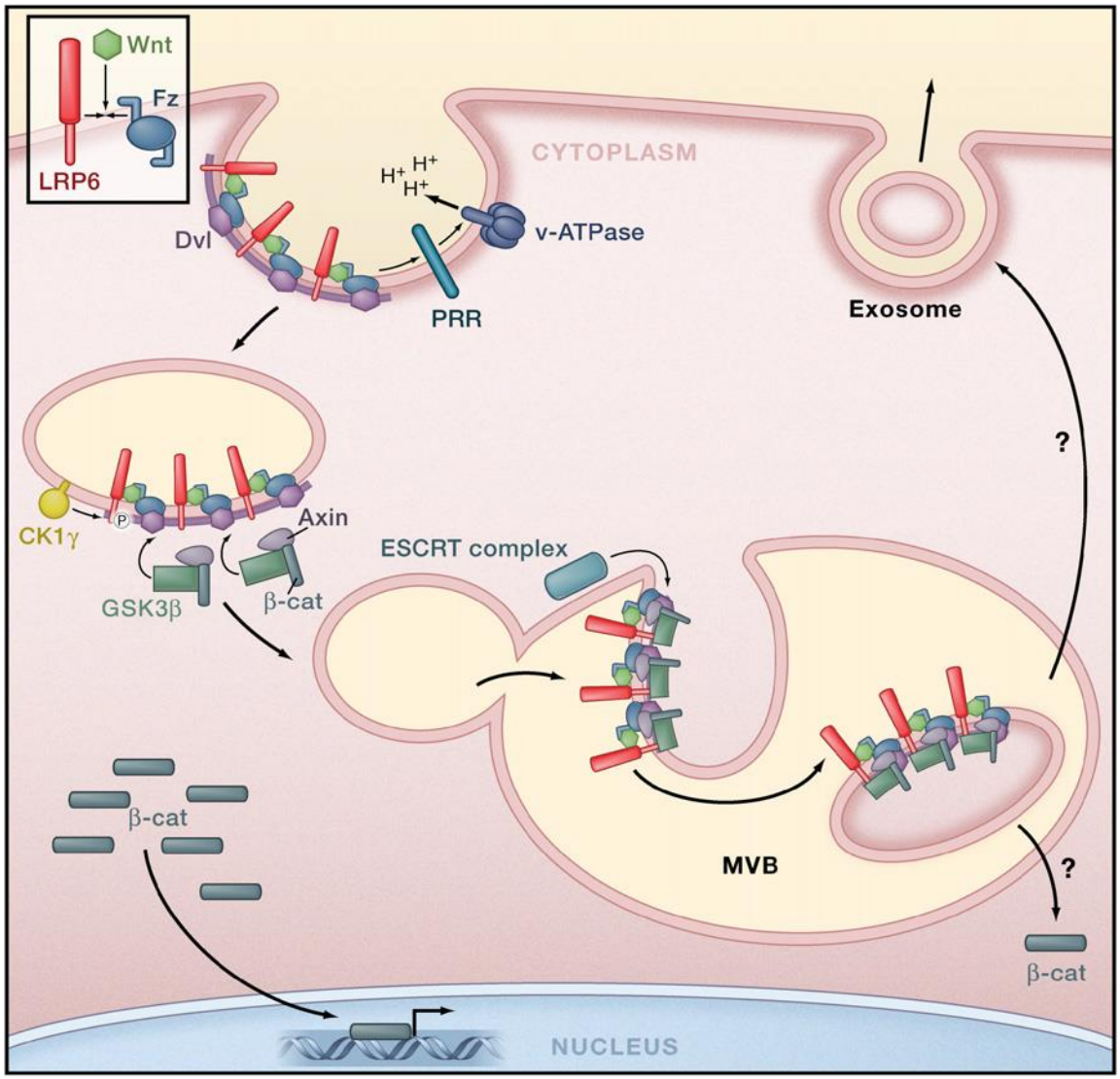


Figure 1-10 – The process of BDC inactivation. Niehrs and Acebron suggest that the correct order of events for is: 1) Wnt ligand binds to Fz/Dvl receptor; 2) Dvl-mediated receptor clustering; 3) First endocytosis and acidification of the lumen; 4) LRP phosphorylation and BDC recruitment; 5) Second endocytosis in the multi-vesicular body. Reproduced from Niehrs and Acebron (2010).

cytosol, preventing β-catenin degradation and thereby allowing it to accumulate (Niehrs and Acebron, 2010; Taelman *et al.*, 2010). However this simplified model is made more complicated by recent findings of Cruciat *et al.* (2010) who reported that endocytosis of the Wnt/Fz/LRP receptor complex occurs at an earlier stage than initially thought. They suggest that acidification of the vesical lumen is required for LRP6 phosphorylation to occur and Axin recruitment to the Wnt/Fz/Dvl receptor complex.

1.4 Leucine-rich Repeat Kinase 2 (LRRK2)

1.4.1 LRRK2 as a Tandem Repeat Protein

LRRK2 is a very large 2527 amino acid, multi-domain protein containing a ROC domain and a kinase domain in the middle, with an eponymous N-terminal LRR domain and a C-terminal WD40 repeat

domain along with suggestions of outer unmapped repeat domains. However the crystal structure of full-length LRRK2 is currently unsolved and there is some debate as to the exact number and locations of these repeats. Cookson (2010) suggests that there are thirteen LRRs, six WD40 repeats and an unspecified number of other repeats, whereas Mills *et al.* (2012) predict via homology modelling fourteen LRRs and seven WD40 repeats as well as thirteen ARM repeats and seven ANK repeats (see Figure 1-11a & b).

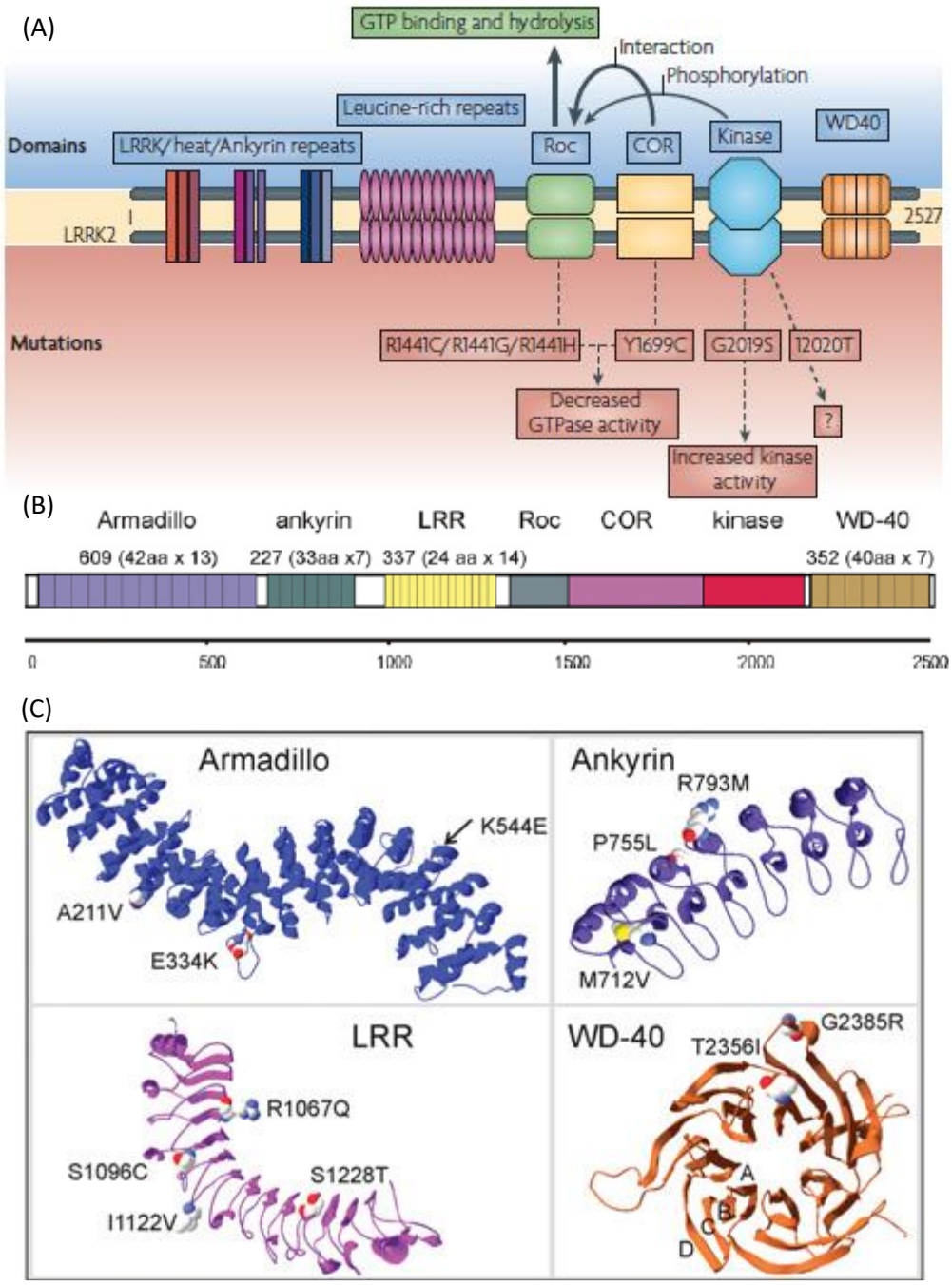


Figure 1-11 – The predicted structural domains of LRRK2. (A) and (B) are the linear arrangement of the domains of LRRK2 as predicted by Cookson (2010) and Mills et al. (2012) respectively, reproduced from their respective papers. (C) is the structures of the repeat domains of LRRK2 as predicted by Mills et al. using homology modelling, with known PD-associated mutations highlighted, also reproduced from their paper.

A significant number of mutations related to Parkinson's disease (PD) have been identified, a number of which have been shown *in vitro* to increase LRRK2's rate of phosphorylation and autophosphorylation (Tong *et al.*, 2010; Yue and Lachenmayer, 2011). Some of these mutations fall outside the catalytic domains, within the proposed repeat domains. These include: A211V, E334K and K544E in the predicted ARM repeat domain; M712V, P755L and R793M in the ANK repeat domain; R1067Q, S1096C, I1122V, and S1228T in the LRR domain; and T2356I and G2385R in the WD40 repeat domain (Mills *et al.*, 2012) (see Figure 1-11c). These proposed tandem repeat domains of LRRK2 are explored further in Chapter 3.

1.4.2 LRRK2 and Parkinson's Disease

Parkinson's disease (PD) is an idiopathic, neurodegenerative disease, resulting from the death of dopaminergic neurones in the substantia nigra region of the midbrain (Dauer and Przedborski, 2003). The main symptoms are tremor at rest, bradykinesia, deteriorating to akinesia, rigidity and postural instability (Jankovic, 2008). In the later stages of the disease, cognitive disturbances develop although they are sometimes observed from the offset. Primarily loss of executive function, sleep disorders, loss of attentiveness and slowed cognitive speed, which can deteriorate into mood and behaviour changes, memory loss and dementia (Caballol, Martí and Tolosa, 2007).

Diagnosis is by neurological examination, however such a test is only 75-90% accurate and a true diagnosis can only be confirmed by autopsy where cell death in the substantia nigra, specifically in the pars compacta, is observed. This region of the brain plays a significant role in motor control (Hodge and Butcher, 1980; Pioli *et al.*, 2008), as well as additional roles in learned responses to stimuli (Ljungberg, Apicella and Schultz, 1992; Da Cunha *et al.*, 2006) and temporal processing (Matell and Meck, 2000) and it has been suggested that it also has a role in the sleep-wake cycle (Lima *et al.*, 2007), accounting for the observed symptoms. The surviving surrounding cells are also observed to have high numbers of inclusions enriched in phosphorylated α-synuclein (Lesage and Brice, 2009) known as Lewy Bodies. The presence of these Lewy Bodies combined with the cell death in the substantia nigra pars compacta is what clinically distinguishes PD from other neurodegenerative diseases such as Dementia with Lewy Bodies.

Currently the only therapies that exist for PD are aimed at treating its symptoms as the underlying causes of the disease are currently unknown. The main treatment is with L-DOPA, a precursor to dopamine, or with dopamine agonists. However their ability to cross the blood-brain barrier is poor so they need to be administered in conjunction with other medications to supress the side-effects, however these may also have their own side-effects. Eventually all patients stop responding to such treatment as all of the dopaminergic neurones die off and there is nothing for them to act upon.

Several genes have been strongly implicated in PD, including α-synuclein (SNCA), microtubule-associated tau protein (MAPT), and LRRK2 (PARK8) in autosomal dominant PD; and E2-ubiquitin ligase parkin (PARK2), PTEN-induced kinase 1 (PARK6), and protein deglycase DJ-1 (PARK7) in recessive forms of PD (Davie, 2008; Drolet, Sanders and Kern, 2011). However Mendelian familial inheritance accounts for only a small percentage of PD cases (around 5-10%, Kang and Marto (2017)), and so it is the convention to refer to Parkinsonism with a genetic cause as "Familial PD", in order to distinguish it from the truly idiopathic "Sporadic PD".

The role of LRRK2 in this disease is not well understood (Aliaga, Cai and Liu, 2012), despite its mutations accounting for 5% of Familial PD cases and 3% of Sporadic PD cases (Tan and Jankovic, 2006), the highest single attributable cause of PD. Primarily this is because LRRK2 is a promiscuous kinase, modulating the activities of many different pathways, thereby acting as a hub and allowing crosstalk. LRRK2 is expressed throughout the brain (Biskup *et al.*, 2006; West *et al.*, 2014) as well as peripheral organs including the kidney, lung, spleen, and peripheral blood mononuclear cells (Su *et al.*, 2004; Larsen and Madsen, 2009). This explains why LRRK2 knockout mice were observed developing deep pathological legions on the kidney and lungs (Herzig *et al.*, 2011).

LRRK2 is primarily a cytoplasmic protein but does associate with membrane organelles such as mitochondria, ER, Golgi apparatus, endosomes and synaptic vesicles (Cookson, 2010) and recent studies have also shown LRRK2 to play a significant role in synaptic transmission (Piccoli *et al.*, 2014). LRRK2 is able to modulate both synaptic vesicle exocytosis, via its ability to phosphorylate Snapin (Piccoli *et al.*, 2011; Yun *et al.*, 2013), and clathrin-mediated endocytosis (Arranz *et al.*, 2015).

This suggestion is further supported by studies that demonstrate that LRRK2 is involved in the dynamic assembly of Filamentous actin (F-actin) and microtubule assembly (Parisiadou and Cai, 2010). LRRK2 phosphorylates the ezrin, radixin and moesin (ERM) family proteins (Jaleel *et al.*, 2007), targeting them to the growing tips of neurites where they tether the F-actin to the cytoplasmic membrane of the protruding filopodia, regulating neurone morphology (Parisiadou *et al.*, 2009). The LRRK2 mutations which increase its phosphorylation rate, also enhances ERM phosphorylation, possibly perturbing the neurone morphology homeostasis (Winner *et al.*, 2011).

LRRK2 also interacts with microtubules through its GTPase domain where it has a stabilising effect (Gandhi *et al.*, 2008; Gillardon, 2009). This has been demonstrated in mouse models, where it was observed that free tubulin was increased in the brain homogenates of LRRK2 knockout mice (Gillardon, 2009), and decreased in those over-expressing LRRK2 or LRRK2 variant with enhanced phosphorylation activity (Lin *et al.*, 2009). Dynamic assembly and disassembly of microtubules is

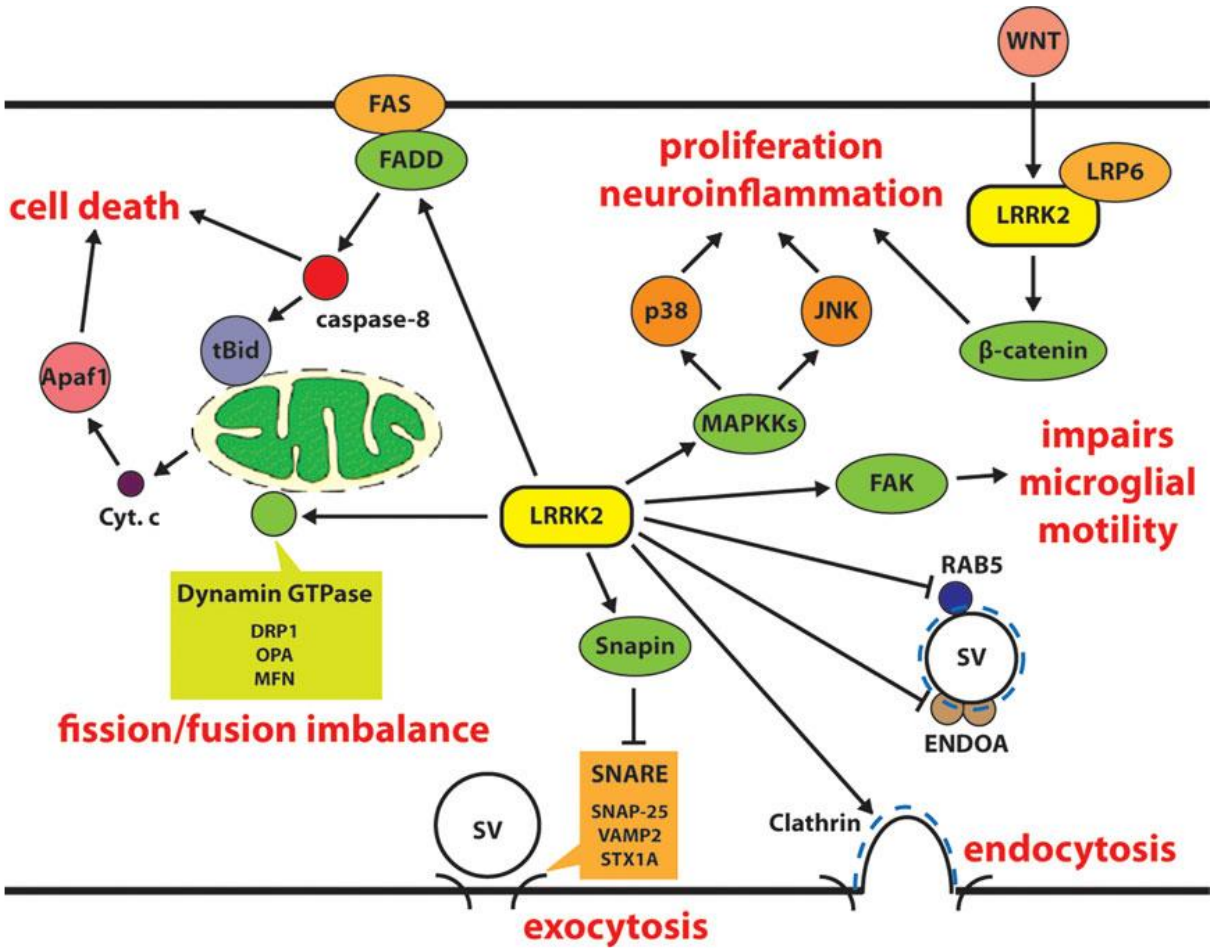


Figure 1-12 – The PD-associated interactions of LRRK2. The proposed interactions of LRRK2 that might lead to neuronal cell dysfunction in Parkinson's disease. Reproduced from Kang and Marto (2017).

important to the health of neurones, so it is thought that this excessive stabilisation may be detrimental. LRRK2 has also been shown to phosphorylate tau protein, a protein that also binds and stabilises microtubules (Drubin and Kirschner, 1986). However phosphorylated tau protein (phospho-tau) is less effective at stabilising microtubules (Johnson and Stoothoff, 2004) which has an opposite effect to LRRK2, suggesting LRRK2 plays a far more complicated role in microtubule regulation than previously understood. Phospho-tau is observed aggregated in neurofibrillary tangles in Alzheimer's patients, which may be a cause of the dementia suffered in late-stage PD (Rizzo *et al.*, 2008).

α-Synuclein (α-syn) is a cytosolic protein whose normal function in neurones is not fully understood but is suggested to play roles in synaptic vesicle dynamics, intracellular trafficking and mitochondrial function as well as potentially being a chaperone (Vekrellis *et al.*, 2011; Wales *et al.*, 2013; Burré, 2015). α-Syn gains its neurotoxicity through aggregation from soluble monomer to soluble oligomers, to small protofibrils and finally large insoluble fibrils. However there is some disagreement as to whether it is the small soluble oligomers (Conway *et al.*, 2000) or the large

insoluble fibrils (Trojanowski and Lee, 1998) that are the neurotoxic species. Phosphorylation at serine-129 seems to be the trigger for this aggregation although the cause of this change in behaviour has not being satisfactorily explained (Iwatsubo, 2003; Sato, Kato and Arawaka, 2013). There is also increasing evidence that aggregated α-syn can propagate in a prion-like manner providing an explanation as to why Lewy Bodies are observed in neurones neighbouring the substantia nigra pars compacta (Brundin, Melki and Kopito, 2010; Brundin and Melki, 2017).

One might speculate that as LRRK2 is a kinase, it is LRRK2 which is responsible for α-syn phosphorylation. However only a single study has been published showing that a fragment reputed to correspond to the catalytic domain of LRRK2 phosphorylates wild-type α-syn to a significant degree (Qing *et al.*, 2009) whereas numerous other studies have failed to replicate these findings. However LRRK2 has been shown to phosphorylate the A53T mutational variant of α-syn at threonine-53, which also had the effect of promoting its rate of aggregation and the progression of the neurodegeneration in transgenic mouse models (Lin *et al.*, 2009).

LRRK2 and α-syn are also believed to mediate cytotoxicity via the 14-3-3 family of proteins (Xu *et al.*, 2002; Nichols *et al.*, 2010). This family of proteins makes up 1% of total brain proteins and are capable of interacting with a variety of protein partners, allowing it to perform multiple cellular functions (Dougherty and Morrison, 2004), modulated by phosphorylation patterns. One important function is that the 14-3-3 proteins sequester proapoptotic proteins (Yuan and Yankner, 2000) and one theoretical cause of cell death is that α-syn binds to and sequesters the 14-3-3, preventing the apoptosis suppression (Recchia *et al.*, 2004). LRRK2 has also been shown to interact with the 14-3-3 family proteins, when the Ser910 and Ser935 are phosphorylated (Dzamko *et al.*, 2010), where it has the potential to alter the phosphorylation state of the 14-3-3 proteins and hence theirs propensity to bind α-syn.

There is increasing evidence that LRRK2 dysfunction might mediate neuron toxicity by disruption of mitochondrial homeostasis (Niu *et al.*, 2012). Dynamin-related protein 1 (DRP1) promotes mitochondrial fission and is activated by phosphorylation at S616. Over-expression of wild-type LRRK2, as well as the PD-associated hyperactive G2019S variant has been shown to enhance DRP1 mediated mitochondrial fragmentation (Wang *et al.*, 2012; Su and Qi, 2013), and hyperphosphorylation of DRP1 at S616 has been observed in non-LRRK2 related sporadic PD (Santos *et al.*, 2015). Similarly the regulators of mitochondrial fusion, optic atrophy 1 and mitofusion 1 and 2, have decreased expression in PD patients carrying the G2019S mutation (Stafa *et al.*, 2014).

Neuroinflammation has also been strongly implicated in the degeneration of dopaminergic neurones in PD (Barnum and Tansey, 2012) for which LRRK2 may play a role (Russo, Bubacco and Greggio, 2014). LRRK2 has been shown to act as a MAPKKK, phophorylating MAPKKs (Gloeckner *et al.*, 2009; Hsu *et al.*, 2010) as well as MAPK scaffolding proteins (Hsu, Chan and Wolozin, 2010) promoting inflamation through activation of c-Jun N-terminal kinases and p38. Indeed MPTP, which is a prodrug to the neurotoxin MPP^+ and is used to produce Parkinsonism in mouse models (Jackson-Lewis and Przedborski, 2007), has been shown to act through both disruption of mitochondrial complex 1 (Kanazawa *et al.*, 2017) and activation of the p38 MAPKs (Hwang *et al.*, 2016).

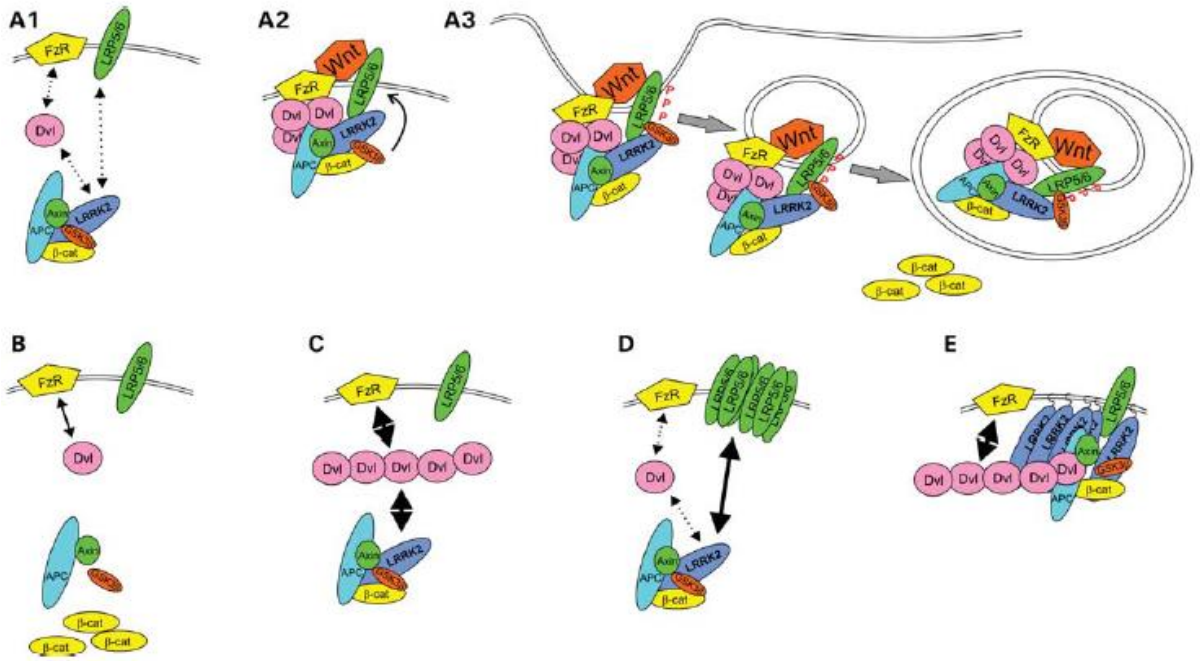


Figure 1-13 – The proposed role of LRRK2 as a canonical Wnt signalling scaffold: Reproduced from Berwick and Harvey (2012). Description taken from that paper: *"(A1) In the basal state, LRRK2 is associated with the BDC. The distribution of BDCs between membrane and cytosol is in dynamic equilibrium dependent on interactions between Fz receptors (FzR) and DVL proteins, between DVL proteins and LRRK2 and between LRP5/LRP6 co-receptors and LRRK2. (A2) Following stimulation with Wnt ligand, DVL proteins are polymerized, thereby recruiting BDCs to the membrane via interaction with LRRK2 and additional BDC proteins. LRRK2 also interacts with the intracellular domain of LRP5/LRP6, thus bridging DVL proteins, the BDC and LRP5/LRP6, and facilitating LRP5/LRP6 phosphorylation by GSK3β, and the formation of LRP5/LRP6 signalosomes. (A3) Signalosomes containing phosphorylated LRP5/LRP6 and GSK3β are internalized into the endosomal system and ultimately sequestered into multi-vesicular bodies. Sequestration of GSK3β prevents phosphorylation of newly synthesized b-catenin allowing the protein to accumulate in the cytoplasm and activate β-catenin-dependent transcription in the nucleus. Whether LRRK2 is sequestered into MVBs with LRP6 and GSK3β or recycled to the cytosol remains to be determined. Note that LRRK2 is involved in multiple steps and interacts with multiple Wnt components. By altering strength of interaction with these proteins, PARK8 mutations are expected to compromise canonical Wnt activity stimulated by Wnt ligands, and perhaps also to alter basal β-catenin activity. (B) The primary effect of loss of LRRK2 is disruption of the BDC, leading to β-catenin stabilization. (C) Overexpressed DVL protein polymerizes and is sufficient to recruit BDCs to the membrane via increased interactions with FzRs and LRRK2. Presumably, the overexpression of LRRK2 further increases the strength of the LRRK2-DVL interaction. (D) Overexpressed LRP6 also increases the membrane recruitment of BDC via binding to LRRK2. (E) Membrane-targeting of LRRK2 ensures the DVL-LRRK2 interaction takes place at membranes, thereby maximizing the membrane recruitment of BDCs."*

1.4.3 LRRK2 in the Canonical Wnt Signalling Pathway

A final way in which LRRK2 has been proposed to play a role in PD, is through modification of the transcription signals from the canonical Wnt signalling pathway (Berwick and Harvey, 2013). As has already been mentioned in Section 1.3, Wnt signalling plays an important role in cell fate determination and homeostasis, including in neurones (Inestrosa and Varela-Nallar, 2015). In particular, it has been shown that Wnt signalling is vital for the development of dopaminergic neurones (Castelo-Branco *et al.*, 2003, 2009; Parish *et al.*, 2008). Recently, Berwick and Harvey (2012) were able to show that LRRK2 is able to associate with key members of the canonical Wnt signalling pathway (Axin, Dvl, GSK3β and β-catenin itself) and that the presence of LRRK2 modulated canonical Wnt signalling activity. They suggest that LRRK2 plays a scaffolding role in the BDC in both its cytosolic and membrane-bound states (see Figure 1-13). While Berwick and Harvey have not mentioned it, I have noted previously in Sections 1.3 and 1.4.2 that LRP6 phosphorylation, an important step in BDC membrane association, has been shown to be mediated by MAPKs (Davidson *et al.*, 2009) and LRRK2 contains a MAPKKK-like domain, suggesting another possible mechanism for LRRK2-based Wnt signal modulation.

1.5 β-Catenin

1.5.1 β-catenin as a Tandem Repeat Protein

β-catenin is the human homologue of the *Drosophila* protein Armadillo, the protein that gives its name to the ARM repeat (Peifer, Berg and Reynolds, 1994), and as such, β-catenin is often used as the main example of an ARM repeat protein. The isolated ARM repeat domain (residues 134-671) of β-catenin was first crystallised in 1997 but it was not until 2008 that structure of the full-length protein was finally solved (Huber, Nelson and Weis, 1997; Xing *et al.*, 2008). This was likely due to the N- and C-terminal regions (1-133 and 672-781 respectively) being mostly disordered, which, as mentioned in Section 1.2, is known to cause problems with crystal packing and formation.

β-catenin itself contains twelve imperfect ARM repeats which stack linearly to form a right-handed superhelix of helices. The third helix (H3) of each ARM repeat lines the groove formed by the superhelix and is enriched in positively charged residues (Huber, Nelson and Weis, 1997), creating an extended docking site for β-catenin's negatively charged binding partners; APC, E-cadherin and the TCF/LEF family of transcription factors (Graham *et al.*, 2000; Eklof Spink, Fridman and Weis, 2001; Poy *et al.*, 2001; Ha *et al.*, 2004; Xing *et al.*, 2004).

1.5.2 β-catenin: function and dysfunction

β-catenin fulfils two important functions *in vivo*, the first of which is that it acts as the signal transducer in the canonical Wnt signalling pathway (see Section 1.3) (MacDonald, Tamai and He, 2009). Upon Wnt stimulation, β-catenin translocates to the nucleus where it binds the TCF/LEF family of transcription factors (Arce, Yokoyama and Waterman, 2006), resulting in the transcription of a number of developmentally important genes including cyclin D1 and c-Myc (Daniels, Eklof Spink and Weis, 2001; Conacci-Sorrell, Zhurinsky and Ben-Ze'ev, 2002). The β-catenin-TCF/LEF complex binds to a DNA consensus sequence called the Wnt Responsive Element, 5'-A-(C/G)-(A/T)-T-C-A-A-A-G-3', causing localised DNA bending (Hatzis *et al.*, 2008). From there, the complex recruits a range of coactivators, including BCL9, Pygo, Mediator, p300/CBP and TRRAP/TIP60 histone acetyltransferases, MLL1/2 histone methyl-transferases, the SWI/SNF family and the PAF1 complex, leading to transcription initiation and elongation and histone and chromatin modification (Willert and Jones, 2006; Mosimann, Hausmann and Basler, 2009).

The other role β-catenin plays *in vivo* is as an adaptor, mediating cell-cell adhesion at adherens junctions. At adherens junctions, β-catenin binds to the intracellular domain of the cadherin family of proteins via β-catenin's positively charged superhelical groove (Huber and Weis, 2001). β-catenin then binds α-catenin via β-catenin's C-terminal domain, and α-catenin binds to the filamentous actin (F-actin) of the cytoskeleton (Pokutta and Weis, 2007). Meanwhile the extracellular domain of the cadherins bind homophilically with cadherins from other cells, thereby bridging the cytoskeletons of neighbouring cells promoting tissue formation (Harris and Tepass, 2010). Adherens junctions effectively compete with the BDC for cytosolic β-catenin (Herzig *et al.*, 2007) and between these two mechanisms, the basal level of cytosolic β-catenin is kept at a safe low level.

From its dual roles in the cell, it can be easily seen that β-catenin is oncogenic and its dysfunction has been implicated in numerous types of cancer (Clevers and Nusse, 2012). In addition to the cell cycle promoters cyclin D1 and c-Myc, Wnt/β-catenin signalling also induces a number of other gene targets known to promote cancer progression, namely matrilysin (Crawford *et al.*, 1999), a pro-metastatic matrix metalloproteinase, and the multidrug resistance protein, Mdr1 (Yamada *et al.*, 2000), so tight control is key (Kim, Kim and Jho, 2013). Mutations in β-catenin, as well as other members of the BDC which help regulate it, are known to cause cancers, primarily in the gut (Korinek *et al.*, 1997; Morin *et al.*, 1997; Liu *et al.*, 2000), with disruption of APC being the single most common event in colorectal cancers (>85%, Bienz and Clevers (2000)).

Additionally, disruption of β-catenin binding at adherens junctions is also cancer promoting (Yoshida *et al.*, 2001). Aside from being inherently pro-metastatic through the weakening of cell-cell adhesion

(Conacci-Sorrell, Zhurinsky and Ben-Ze'ev, 2002), adherens junction-associated β-catenin represents a significant cellular pool of β-catenin (Herzig *et al.*, 2007), the release of which can result in nuclear accumulation and unregulated gene activation (Polakis, 2007). Studies have shown that loss of E-cadherin in particular is implicated in the transformation of benign adenoma cells to malignant carcinoma cells and that its reintroduction reduced cancel cell metastasis (Conacci-Sorrell, Zhurinsky and Ben-Ze'ev, 2002; Gail, Frank and Wittinghofer, 2005).

β-catenin would therefore seem like a prime target for intervention in cancer, however it has so far proven to be a difficult goal. The challenge lies in the binding interface of the ARM superhelical groove for which many of its binding partners compete (see Figure 1-14). Any drug aimed at modulating one of the functions of β-catenin needs to be carefully designed so as not to disrupt any of its other functions.

Table 1-4: Binding affinities of β-catenin binding to its partners.

Binding Partner	K_d (nM)*
ICAT[1]	12.0 ± 1.8
APC repeat 3[1]	3100 ± 300
phosphorylated APC repeat 3[1]	10 ± 1.3
TCF7L2[2]	16 ± 3
E-cadherin cytoplasmic domain (E_{cyto})[1]	46 ± 4.4
phosphorylated E_{cyto}[1]	0.052 ± 0.005
Axin[1]	1300 ± 200
Lef-1 (1-65)[1]	49 ± 3.6

* As determined by ITC. Exact conditions varied between studies

[1] Choi, Huber and Weis (2006), [2] Sun and Weis (2011)

That being said, there have been a number of successful attempts at directly targeting β-catenin made in recent years. Mani *et al.* (2009) were able to reduce the expression levels of several Wnt-induced cell cycle regulators in mouse xenograft models using short hairpin RNA (shRNA) to target and disrupt the interaction between β-catenin and its coactivator BCL9. Then later Takada *et al.* (2012) were able to replicate this same effect in mouse xenograft models using a hydrocarbon-stapled α-helical peptide based on the structure of BCL9. However it should be noted that BCL9 binds to a different region of β-catenin (the N-terminal-most ARM repeats) not to the superhelical groove (de la Roche *et al.*, 2012).

Grossmann *et al.* (2012) took a similar approach to Takada *et al.* (2012) in inhibiting Wnt signalling. They designed a hydrocarbon-stapled α-helical peptide based on the sequence and structure of the α-helix formed when Axin binds to β-catenin, which does bind in the superhelical groove. The intention was to compete directly with Tcf4 for the superhelical groove and thereby inhibit transcription, and they were able to inhibit cell proliferation in cell lines containing known mutations

of APC and β-catenin that prevent β-catenin ubiquitination and degradation. They also designed and tested peptides based on the Tcf4 sequence but found at an early stage that an equivalently sized peptide based on Tcf4 did not exhibit significant binding to β-catenin.

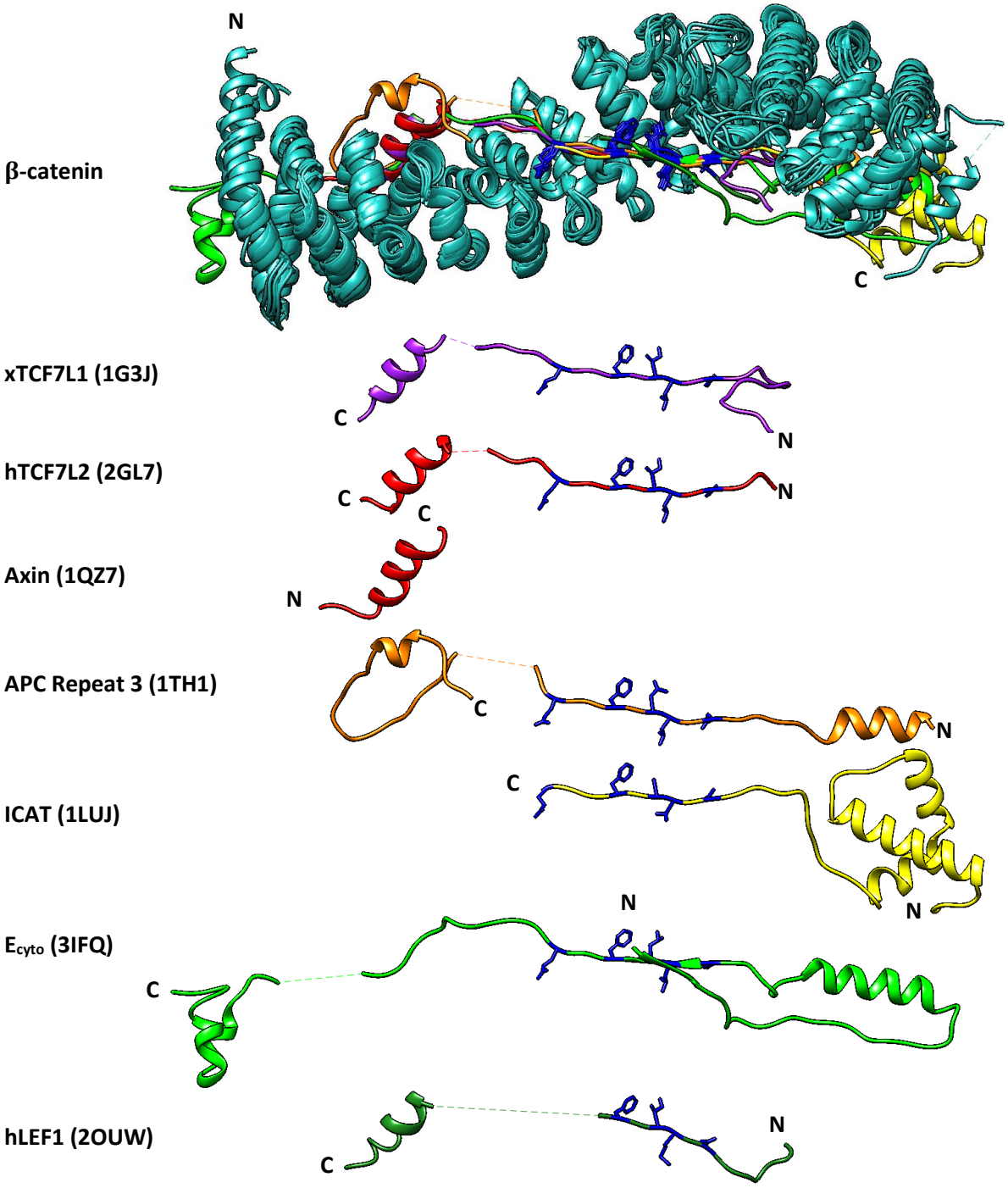


Figure 1-14 – β-catenin binding to its partners: A superposition of the crystal structures of β-catenin (cyan) in complex with its binding partners (purple through to dark green), visualised in UCSF Chimera: *Xenopus* TCF7L1 (PDB ID 1G3J, Graham *et al.* (2000)), human Tcf4 (2GL7, Sampietro *et al.* (2006)), Axin (1QZ7, Xing *et al.*, (2003)), APC third repeat (1TH1, Xing *et al.* (2004)), ICAT (1LUJ, Graham *et al.* (2002)), the cytoplasmic domain of E-cadherin (3IFQ, Choi *et al* (2009)) and human Lef-1 (3OUW, Sun and Weis (2011)). Some of their individual affinities are listed in Table 1-4. α-helices are represented by cylinders and unstructured regions are denoted by a dashed line. All partners except axin, bind antiparallel to β-catenin and the consensus binding motif, $Dx\Phi\Phi x\Omega x_{2\text{-}7}E$ is highlighted (dark blue) (where Φ indicates a hydrophobic residue and Ω indicates an aromatic residue).

1.6 TCF7L2

Transcription factor 7-like 2 (TCF7L2), also known as Tcf4, is a member of the T-cell factor/lymphoid enhancer-binding factor (TCF/LEF) family of transcription factors. In mammals, there are four members of the TCF/LEF family: Transcription factor 7 (TCF7, also known as Tcf1), Lymphoid enhancer-binding factor 1 (LEF1), Transcription factor 7-like 1 (TCF7L1, also known as Tcf3) and TCF7L2, each of which contains a conserved N-terminal β-catenin binding domain (Behrens *et al.*, 1996), a DNA binding high mobility group (HMG) box domain, which recognises the Wnt responsive element (Hatzis *et al.*, 2008) and a nuclear localisation signal. The crystal structures of the TCF/LEF family have yet to be solved but it is known that the N-terminal β-catenin binding domain is intrinsically disordered, gaining some structure on binding (Graham *et al.*, 2000, 2001; Poy *et al.*, 2001; Sampietro *et al.*, 2006).

In the absence of the Wnt signal, the TCF/LEF family members bind to the Groucho/TLE family of corepressors to inhibit the transcription of the Wnt target genes (Brantjes *et al.*, 2001). It is interesting to note that, mirroring β-catenin (see Section 1.5.2), Groucho actually acts through the recruitment of further corepressors, namely histone deacetylases 1 & 2 (HDAC1 & 2) (Chen *et al.*, 1999; Brantjes *et al.*, 2001). Then on canonical Wnt activation, Groucho/TLE is displaced from the TCF/LEFs by β-catenin leading to transcription of those genes (MacDonald, Tamai and He, 2009). The TCF/LEF family members are not functionally redundant, with LEF1 behaving as a transcriptional activator, TCF7L1 as a transcriptional repressor, and TCF7 and TCF7L2 behaving in a context-dependant manner (Cadigan and Waterman, 2012).

Of the four TCF/LEF family members, TCF7L2 has been the most widely studied and is also the most ubiquitously expressed in adult tissues (Archbold *et al.*, 2012; Hrckulak *et al.*, 2016). TCF7L2 is most highly expressed in the small intestine and colon where it plays a vital role in proper development of the small intestine and in the homeostasis of the colonic epithelium in adult tissue (van Es *et al.*, 2012). As a result, TCF7L2 mutations are implicated in many types of cancers (Tang *et al.*, 2008) and colorectal cancers in particular (Hazra *et al.*, 2008; Slattery *et al.*, 2008). Single nucleotide polymorphisms within the TCF7L2 gene are also a genetic biomarker for an increased risk for the development of both type 2 and gestational diabetes (Vaquero *et al.*, 2012; Zhang *et al.*, 2013). This is thought to be due to the aberrant activation of Wnt target genes, which repress proglucagon synthesis (Jin and Liu, 2008).

1.7 Project Aims

Understanding the intricacies of the canonical Wnt signalling pathway is one of the biggest challenges in science today. The pathway is implicated in a range of human diseases which have so far proven to be resistant to treatment including Parkinson's disease, diabetes mellitus, and a range of different cancer types. The aim of this project is to broaden our understanding of the interplay between some of the members of the canonical Wnt signalling pathway, namely the Parkinson's disease-implicated, repeat protein LRRK2, and intrinsically disordered transcription factor TCF7L2. The biophysical data generated in this study are intended to further our understanding of the canonical Wnt signalling pathway and to support the rational design of novel therapeutics in the treatment of the diseases associated with this pathway.

In Chapter 3, I discuss experiments to develop a protocol for the expression, purification and storage of the previously unmapped protein-protein interaction domains of the large, multi-domain protein leucine-rich repeat kinase 2 (LRRK2). These purported domains have recently, been predicted to be comprised of tandem repeats of the ANK, ARM, LRR and WD40 families (Mills *et al.*, 2012). The intension was to biophysically characterise the isolated domains: determine their equilibrium stabilities to chemical and thermal denaturation using circular dichroism and intrinsic fluorescence as well as their kinetic folding and unfolding rates via stopped-flow fluorescence and circular dichroism, as well as attempting to obtain crystal structures to confirm their identities.

Chapter 4 focuses on the development of a protocol for the expression and purification of the N-terminal β-catenin-binding domain of wild-type TCF7L2. The equilibrium binding constant of the wild-type TCF7L2 for β-catenin will be determined using isothermal titration calorimetry (ITC) and the kinetic rate constants for association and dissociation determined by stopped-flow fluorescence spectroscopy. A fluorescent reporter system will need to be developed as the intrinsic fluorescence of β-catenin and fluorescently-labelled β-catenin have both historically been inadequate reporters of partner binding.

The work described in Chapter 5 focuses on understanding the driving forces that allow intrinsically disordered TCF7L2 to display three distinct structures when bound to β-catenin. A series of variant constructs to alanine residues thought to be key to TCF7L2-β-catenin binding, but which vary between structures, will be created, and their effect on the association and dissociation kinetics determined using the methods described in Chapter 4.

In Chapter 6, I map the kinetic pathway of complex formation through a series of TCF7L2 variants within the regions in TCF7L2 that are conserved between the three crystal structures. These variants

will be used investigate the potential for multiple association and dissociation pathways, analogous to the multiple folding and unfolding pathways observed for tandem repeat proteins. The effect of these mutations on the kinetic rate constants are determined by a combination of stopped-flow fluorescence spectroscopy and time-dependent fluorescence spectroscopy.

Finally, the work in Chapter 7 covers other aspects of the TCF7L2-β-catenin interaction. The role of electrostatic interactions in TCF7L2-β-catenin binding is investigated in the wild-type protein and a variant with a lower net charge. The effect of increasing electrostatic shielding on the kinetic rate constants for both constructs is determined by the methods described in Chapter 6. To determine whether TCF7L2-β-catenin binding follows an induced folding or a conformational selection pathway stapled α-helical TCF7L2 constructs are made. And to determine whether TCF7L2-β-catenin binding follows an avidity, allovalency or fuzzy complex binding mechanism, truncated TCF7L2 constructs are made.

Chapter 2 - Materials and Methods

2.1 Reagents and Solutions

All chemical reagents were purchased from Sigma-Aldrich Chemical Company unless otherwise stated.

2.1.1 *E. coli* bacterial strains

All cell lines are divided into 50 µl aliquots, flash frozen with liquid nitrogen and stored at -80°C until use:

- α-Select Gold Competent Cells, which are a DH5α derivative modified for increased transformation efficiency and recombinant plasmid production, were purchased from Bioline.
- α-Select Bronze Competent Cells, which are the same derivative as α-Select Gold Competent Cells but lower colony forming efficiency, were also purchased and used to create a lab stock culture.
- C41 Competent Cells (Miroux and Walker, 1996), which are a BL21(DE3) derivative modified for increased transformation efficiency, recombinant protein overexpression and toxic protein resistance, were from our lab stock culture.
- B834(DE3) Competent Cells (Wood, 1966), which are the methionine auxotrophic parental strain of BL21(DE3) and allow for high specificity labelling of target proteins at methionine residues e.g. ^{35}S-methionine or selenomethionine for crystallography (Leahy *et al.*, 1992), were purchased from Novagen.

2.1.2 Plasmids and DNA Primers

Custom primers were ordered from either Integrated DNA Technologies or Sigma Aldrich and are listed in Appendix B. The plasmids used in this Thesis are detailed below:

- **pUC57** - pUC57 is at iterative derivative of pUC19 (Yanisch-Perron, Vieira and Messing, 1985) which contains the ampicillin (amp) resistance gene, bla(Ap^R) and no tag. This vector was not used for expression and the genes supplied in this vector were subcloned into other plasmids used for expression.

- **pGEX** - contains amp resistance and a thrombin cleavable (Leu-Val-Pro-Arg-|-Gly-Ser) glutathione S-transferase (GST) fusion tag. Expression is under the T7 bacteriophage promoter and the lacI repressor which allows protein expression to be induced in the presence of isopropyl β-D-1 thiogalactopyranoside (IPTG).

- **pRSET-A** - contains amp resistance and a thrombin cleavable N-terminal Hexa-His tag (originally purchased from Life Technologies). Expression is under the T7/lacI system.

- **pRSET-GST** which contains amp resistance and a thrombin cleavable N-terminal GST-fusion tag. This is a custom plasmid used by our lab, it is a derivative of pRSET-A (described above) where the Hexa-His tag has been removed and replaced with a GST-fusion tag transferred from pGEX-2T (originally from Pharmacia). Expression is under the T7/lacI system.

- **pET-41a(+)** - contains kanamycin (kan) resistance, a thrombin cleavable combined N-terminal GST-fusion tag followed by a Hexa-His tag (thrombin cleaves both) followed by an enterokinase cleavable (Asp-Asp-Asp-Asp-Lys-|-X) N-terminal S-tag (enterokinase would cleave all three), and a C-terminal Octa-His tag (originally purchased from Merck Millipore). Expression is under the T7/lacI system.

- **pACYCDuet-1** - contains chloramphenicol resistance and two multiple cloning sites (MCSs). The first MCS contains an un-cleavable N-terminal Hexa-His tag, whilst the second MCS contains an un-cleavable C-terminal S-tag (originally purchased from Merck Millipore). Expression is under the T7/lacI system.

2.1.3 Buffers for Molecular Biology

- 50x TAE Buffer (Stock) – 2 M Tris base, 1 M acetic acid, 50 mM EDTA. TAE buffer solubilises DNA by sequestering metal cations, exposing the negatively charged phosphate backbone.

- 10000x SYBR® Safe – purchased from ThermoFisher Scientific. SYBR Safe is a safer alternative to ethidium bromide, it also chelates with DNA allowing it to be visualised under UV-light but is much less cell penetrating.

- Gel Solubilisation buffer – 20 mM Tris-HCl pH 6.6, 5.5 M guanidinium-SCN

- Resuspension buffer – 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNaseA. RNaseA is not stable in Lysis buffer and so is present in the Resuspension buffer and is stored at 4°C

- Lysis buffer – 200 mM NaOH , 1% (w/v) SDS

- Neutralisation buffer – 4.2 M Guanidinium-HCl, 0.9 M KOAc, pH 4.8 with AcOH
- Wash buffer – 10 mM Tris-HCl pH 7.5, 80% (v/v) EtOH
- Elution buffer – 10 mM Tris-HCl pH 7.5

2.1.4 Growth Media and Buffers for Protein Expression

- 2xYT culture medium containing 1.6% (w/v) Tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl was purchased from Formedium.
- SelenoMet™ Medium. M9 minimal media supplemented with glucose, vitamins and all amino acids except methionine, so as to allow proteins to be labelled with selenomethionine or ^{35}S-methionine was purchased from Molecular Dimensions Ltd.
- LB (*amp*$^{+}$) agar plate containing: 1.5% (w/v) agar, 1.0% (w/v) Tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 50 µg ml^{-1} ampicillin; and LB (*kan*$^{+}$) agar plate containing: 1.5% (w/v) agar, 1.0% (w/v) Tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 100 µg ml^{-1} kanamycin were provided by the MRC Laboratory of Molecular Biology.
- 50 mg ml^{-1} ampicillin solution (amp). This is a 1000x concentrated stock, frozen and stored at -20°C.
- 50 mg ml^{-1} kanamycin solution (kan). This is a 1000x concentrated stock, frozen and stored at -20°C.
- 34 mg ml^{-1} chloramphenicol in 100% ethanol. This is a 1000x concentrated stock, stored at -20°C.
- LB (amp^{+}, kan^{+}) agar plates containing: 32 g l^{-1} Lennox L Agar (Invitrogen), 50 µg ml^{-1} ampicillin, 50 µg ml^{-1} kanamycin; LB (amp^{+}, chloramphenicol^{+}) agar plates containing: 32 g l^{-1} Lennox L Agar, 50 µg ml^{-1} ampicillin, 34 µg ml^{-1} chloramphenicol; LB (kan^{+}, chloramphenicol^{+}) agar plates containing: 32 g l^{-1} Lennox L Agar, 50 µg ml^{-1} kanamycin, 34 µg ml^{-1} chloramphenicol; and LB (amp^{+}, kan^{+}, chloramphenicol^{+}) agar plates containing: 32 g l^{-1} Lennox L Agar, 50 µg ml^{-1} ampicillin, 50 µg ml^{-1} kanamycin, chloramphenicol 34 µg ml^{-1} chloramphenicol.
- 1 M isopropyl β-D-1 thiogalactopyranoside solution (Stock) (IPTG). This stock is frozen and stored at -20°C.

2.1.5 Buffers for Protein Purification

Buffers containing DTT or protease inhibitors were made fresh from stock solutions on the day of use.

Stock solutions

- 10x PBS pH 7.4 solution – 80 g/l NaCl, 2 g/l KCl, 14.4 g/l Na_2HPO_4, 2.4 g/l KH_2PO_4 buffered to pH 7.4 with KOH. Sterile Filtered.
- 1 M Tris-HCl at pHs 7.5, 8.5 and 8.9
- 1 M dithiothreitol solution (DTT). This stock is frozen and stored at -20°C.
- 1 M imidazole-HCl pH 7.4. Imidazole is a weak base and will affect the pH of 1x PBS pH 7.4, so the 1M imidazole solution was acidified with HCl until pH 7.4.
- 5 M NaCl solution.

His-tagged LRRK2 Constructs

- Lysis buffer – 1x PBS pH 7.4, EDTA-free Protease Inhibitor Cocktail Tablets (Sigma) (1 tablet/50 ml).
- Wash buffer – 1x PBS pH 7.4, 2 mM DTT, 20 mM imidazole-HCl pH 7.4.
- Elution buffer – 1x PBS pH 7.4, 2 mM DTT, 200 mM imidazole-HCl pH 7.4.
- G75 buffer – 50 mM Tris-HCl pH 7.4, 150 mM NaCl.

GST-tagged LRRK2 Constructs

- Lysis buffer – 50 mM Tris-HCl pH 7.5, EDTA-free Protease Inhibitor Cocktail Tablets (Sigma) (1 tablet/50 ml).
- High Salt Wash buffer – 50 mM Tris-HCl pH 7.5, 2 mM DTT, 1 M NaCl
- Low Salt Wash buffer – 50 mM Tris-HCl pH 7.5, 2 mM DTT, 150 mM NaCl
- G75 buffer – 50 mM Tris-HCl pH 7.4, 150 mM NaCl.

His-tagged TCF7L2 Constructs and variants

- Lysis buffer – 1x PBS pH 7.4, EDTA-free Protease Inhibitor Cocktail Tablets (1 tablet/50 ml), 1 mM DTT.
- Wash buffer – 1x PBS pH 7.4, 1 mM DTT, 50 mM imidazole-HCl pH 7.4.
- Elution buffer – 1x PBS pH 7.4, 1 mM DTT, 200 mM imidazole-HCL pH 7.4.
- G75 buffer – 1x PBS pH 7.4, 1 mM DTT.

GST-tagged TCF7L2 Constructs and variants

- Lysis buffer – 50 mM Tris-HCl pH 7.5, EDTA-free Protease Inhibitor Cocktail Tablets (Sigma) (1 tablet/50 ml), 1 mM DTT, 1 M NaCl.
- High Salt Wash buffer – 50 mM Tris-HCl pH 7.5, 1 mM DTT, 1 M NaCl.
- Low Salt Wash buffer – 50 mM Tris-HCl pH 7.5, 1 mM DTT, 50 mM NaCl.
- G75 buffer – 50 mM Tris-HCl pH 7.5, 1 mM DTT, 150 mM NaCl.

GST-tagged β-catenin Weis Construct

- Lysis buffer – 50 mM Tris-HCl pH 8.5, EDTA-free Protease Inhibitor Cocktail Tablets (Sigma) (1 tablet/50 ml), 1 mM DTT, 1 M NaCl.
- High Salt Wash buffer – 50 mM Tris-HCl pH 8.5, 1 mM DTT, 1 M NaCl.
- Low Salt Wash buffer – 50 mM Tris-HCl pH 8.9, 1 mM DTT, 50 mM NaCl.
- G75 running buffer – 50 mM Tris-HCl pH 8.9, 1 mM DTT, 150 mM NaCl.
- High Salt MonoQ buffer – 50 mM Tris-HCl pH 8.9, 1 mM DTT, 1 M NaCl.
- Low Salt MonoQ buffer – As Low Salt Wash buffer.

Buffers for SDS-PAGE

- 20x NuPAGE® MES-SDS Running Buffer (Novagen)
- 20x RunBlue® MES-SDS Running Buffer (Expedeon)
- 20x RunBlue® Teo-Tricine Running Buffer (Expedeon)

- Coomassie Blue Stain – 60-80 mg Coomassie Brilliant Blue G-250 in 1 l mQ water, acidified to approximately pH 2 with 3 ml concentrated HCl (Lawrence and Besir, 2009)

2.2 Molecular Biology

2.2.1 Gift Plasmids

Plasmids encoding the armadillo (ARM) repeat domain of human β-catenin (residues 134-671) were given to us as a kind gift from Professor W.I. Weis. The human β-catenin (residues 134-671) construct (hereafter referred to as the Weis construct) is in a pGEX derived vector which contains ampicillin (amp) resistance and an N-terminal, thrombin cleavable (Leu-Val-Pro-Arg-|-Gly-Ser), glutathione S-transferase (GST) fusion tag. Expression is under the tac promoter, a synthetic promoter which is a hybrid of the trp and lac promoters, and the lacI repressor which allows protein expression to be induced in the presence of IPTG (de Boer, Comstock and Vasser, 1983).

2.2.2 Subcloning using Restriction Endonucleases

Synthetic genes for human LRRK2 subdomains ARM, ANK, LRR and WD40 were ordered from GenScript USA Inc.'s Gene Synthesis Service and were supplied in the pUC57 plasmid vector. pUC57 is at iterative derivative of pUC19 (Yanisch-Perron, Vieira and Messing, 1985) which contains the amp resistance gene, bla(Ap^R) and no tag. This vector was not a useful vector for *in vitro* protein studies due to this lack of a purification tag so these constructs required subcloning into other plasmids more useful for expression. Human TCF7L2 (residues 1-54) was short enough to be ordered as a gene strings and similarly required subcloning into an expression plasmid. It and all other gene strings were ordered from ThermoFisher Scientific's GeneArt Gene Synthesis Service. All of these constructs were designed to have an N-terminal BamHI restriction site (GGATCC) and C-terminal HindIII restriction site (AAGCTT), two restriction sites commonly found in MCSs, for this reason.

A double digest was set up on ice as described in Table 2-1 below, and incubated at 37°C overnight. The reactions were then heated to 80°C for 10 min to inactivate both BamHI and HindIII and the results were analysed by agarose gel electrophoresis and the correct products were purified by gel extraction, as described in Section 2.2.3. Separate digestions were set up for each of my four LRRK2 constructs and for TCF7L2 which was done 12 months later.

Although the cut ends of the recipient plasmid are incompatible they can still associate through non-covalent interactions, which can drastically reduce ligation efficiency. This can be reduced by dephosphorylating the 5' ends of the cut plasmid; this also has the effect that the plasmid could not

religate even if it did have compatible ends. Digest B was taken neat and a reaction was set up with Antarctic Phosphatase (Rina *et al.*, 2000) (NEB UK) on ice as described in Table 2-2 below. The reaction mixture was incubated at 37°C for 1 hr before being heat inactivated by heating to 70°C for 5 min. The dephosphorylated plasmid DNA can be used immediately for ligation or stored at -20°C for future use.

Table 2-1: Digestion conditions for subcloning the LRRK2 constructs

Digest A – LRRK2 Constructs		Digest B – Recipient Plasmid		Digest C – Gene Strings	
2 µl	10x FastDigest Green Buffer	2 µl	10x FastDigest Green Buffer	2 µl	10x FastDigest Green Buffer
10 µl	Construct in pUC57 at 100 ng μl^{-1}	Up to 1 µg	pRSET/pRSET-GST	10 µl	Gene string at 50 ng μl^{-1}
1 µl	BamHI-FastDigest (1 unit*)	1 µl	BamHI-FastDigest (1 unit*)	1 µl	BamHI-FastDigest (1 unit*)
1 µl	HindIII-FastDigest (1 unit*)	1 µl	HindIII-FastDigest (1 unit*)	1 µl	HindIII-FastDigest (1 unit*)
6 µl	mQ water	Up to 16 µl	mQ water	6 µl	mQ water
20 µl	Total volume	20 µl	Total volume	20 µl	Total volume
1 µg	Total DNA	1 µg	Total DNA	0.5 µg	Total DNA

* 1 unit is defined as sufficient to cleave 1 µg of lambda DNA in 5 minutes at 37°C in 1X FastDigest Buffer

Table 2-2: Reaction conditions for dephosphorylating recipient vector

Dephosphorylation of Plasmid DNA	
20 µl	Digest B (neat*)
2.5 µl	10x Antarctic Phosphatase reaction buffer
2.5 µl	Antarctic Phosphatase (12.5 units**)
25 µl	Total Volume

*NEB state that their Antarctic phosphatase can be added to any restriction reaction under any conditions

**1 unit is defined as being the amount of enzyme that will dephosphorylate 1 µg of pUC19 vector DNA cut with HindIII (5´ protruding ends), HincII (blunts ends) or PstI (5´ recessed ends) in 30 minutes at 37°C. Dephosphorylation is defined as > 95% inhibition of recirculation in a self-ligation reaction and is measured by transformation into E. coli.

The purified DNA fragments corresponding to my LRRK2 constructs from digests A and the cut and dephosphorylated recipient plasmid were ligated using T4 Ligase (Murray, Bruce and Murray, 1979) (NEB UK) as described in Table 2-3 below. The construct DNA and plasmid DNA were pre-mixed and heated to 50°C for 10 min to discourage self-association/recircularisation, and then cooled on ice to encourage the constructs and plasmid to associate before adding buffer and ligase. Although the "ideal" vector to insert molar ratio is 1:3, the concentration of construct DNA was undetermined so I elected to saturate the plasmid with excess construct DNA. The reaction mixture was incubated at 16°C overnight and for most of the following day before being transformed into α-Select Gold cells as described in Section 2.2.4 and miniprepped as described in Section 2.2.5.

Table 2-3: Ligation conditions for subcloning the LRRK2 constructs using T4 ligase

	Ligation using T4 Ligase
1 µl	10x T4 DNA Ligase buffer
1 µl	Recipient plasmid DNA
7 µl	Construct DNA
1 µl	T4 DNA Ligase (400 units*, double manufacture suggested concentration)
10 µl	Total volume

*One unit is defined as the amount of enzyme required to give 50% ligation of HindIII fragments of λ DNA (5´ DNA termini concentration of 0.12 µM, 300- µg/ml) in a total reaction volume of 20 µl in 30 minutes at 16°C in 1X T4 DNA Ligase Reaction Buffer.

Ligations of the LRRK2 extensions and the TCF7L2 gene string were performed using QuickStick Ligase (a derivative of T4 Ligase) (Bioline). The reaction condition are described in Table 2-4 below and the method is identical except the reaction mixture only needed to be incubated for 15 min at room temperature.

Table 2-4: Ligation conditions for subcloning the TCF7L2 constructs using QS ligase

	Ligation using QuickStick Ligase
5 µl	4x QS Buffer
Up to 100 ng	Total DNA
1 µl	QuickStick Ligase (1 unit*)
Up to 14 µl	mQ water
20 µl	Total volume

*1 unit is defined as sufficient to ligate 99% of λ/HindIII cohesive-ended fragments, or 80% of λ/EcoRV blunt-ended fragments, in 5 minutes at room temperature. 100% ligation of blunt-ended fragments can be achieved by increasing the ligation time to 15 minutes at room temperature.

2.2.3 Agarose Gel Electrophoresis and Gel Extraction

Different sized DNA molecules were separated by electrophoresis through an agarose-TAE gel. A gel consisting of 1xTAE buffer, 0.5-2% (w/v) agarose and 1x SYBR® Safe was cast in a mold and ran at 100V for 30 min using 1xTAE buffer as a source of conductance. The percentage (w/v) of agarose was varied to optimise separation of the desired product from contaminants: for DNA molecules between 1 and 5 kilobases, 1% agarose was used; for molecules less than 1 kilobase, 2% agarose was used; and for molecules greater than 5 kilobases, 0.7% agarose was used. DNA was visualised by stimulation with UV-light, causing the SYBR® Safe to fluoresce in the presence of nucleobases, and photographed using an Alpha Imager (Alpha Innotech Corporation).

Bands corresponding to DNA of the desired length were excised and extracted from the agarose gel by a method derived from Hamaguchi and Geiduschek (1962) and Vogelstein and Gillespie (1979). Premade kits are available for this (e.g. Qiagen Gel Extraction kit) but our lab prefers to use our own

buffer recipes, detailed in Section 2.1.3, and EZ-10 spin columns (purchased from Bio Basic Canada Inc.). The protocol is otherwise identical to that of the Qiagen Gel Extraction kit.

2.2.4 Transformation of Competent Cells

50 µl of cells were thawed on ice then 50-500 ng of plasmid (depending on purpose and cell line) was added and incubated on ice for 5-30 min. The cell line used depended on the purpose: C41 was used for the majority of protein expression, B834 cells were used for the expression of proteins incorporating unnatural amino acids, α-Select Bronze for miniprepping a known plasmid (for increasing stocks) and α-Select Gold for miniprepping a newly ligated plasmid where reliability is key. The tubes were then heat shocked by heating to 42°C for 45 s in a water bath and then returned to ice for a further 2 min. 200 µl of fresh 2xTY medium without antibiotics was added to the aliquot and the tube was incubated in a shaking incubator at 37°C at 550 rpm and allowed to grow out for 1 hr. The transformed cells were then plated out on an LB agar plate containing the appropriate antibiotic and incubated in a static incubator at 37°C overnight.

2.2.5 Plasmid Amplification and Purification by Miniprep

α-Select Gold or Bronze-efficiency competent cells are *E. coli* strains optimised for plasmid production used by our lab for creating new plasmids stocks or amplifying existing ones respectively. After transforming as described in Section 2.2.4, the plates were then refrigerated at 4°C until later that day when individual colonies (3 or 4 per plate) were picked and inoculated into 5 ml of 2xTY media containing the appropriate antibiotics and grown overnight at 37°C in a shaking incubator.

The cells from the overnight culture were harvested by centrifugation at 1000 g for 10 min at 4°C, after which the supernatant was discarded. Plasmids were extracted and purified from the cell pellet using a method of plasmid purification derived from Birnboim and Doly (1979) and Vogelstein and Gillespie (1979). Premade kits are available for this (e.g. Qiagen Miniprep kit) but our lab prefers to use our own buffer recipes, detailed in Section 2.1.3, and EZ-10 spin columns (purchased from Bio Basic Canada Inc.). The protocol is otherwise identical to that of the Qiagen Miniprep kit.

The purified plasmids concentration and purity was determined by spectrophotometry using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific) and it sequence determined by Eurofins Genomics Sanger sequencing service.

2.2.6 Site-Directed Mutagenesis (SDM)

Matched ssDNA oligonucleotide primer pairs, each with the desired mutation encoded within them, were designed by visual inspection of the DNA sequence determined as described in Section 2.2.5.

Primers were designed to ideally have a melting temperature (T_m) 60- 72°C as determined by NEB's T_m calculator. For the Phusion™ High-Fidelity DNA Polymerase (Phusion), T_m is calculated using the formula described by Breslauer *et al.*, (1986) corrected for different salt concentrations using the Schildkraut formulae detailed in Owczarzy *et al.*, (2004). For Q5® High-Fidelity DNA Polymerase (Q5), T_m is instead calculated using the formulae described in (SantaLucia, 1998), similarly corrected for salt concentrations. The primers were also designed to ideally have a GC content 50-60% and an overlap of at least 10 bases. All the primers used are listed in Appendix A.

Site-Directed Mutagenesis (SDM) was performed by Polymerase Chain Reaction (PCR) to incorporate and amplify mutations encoded by the primer pairs. Reactions were carried out using either Phusion or Q5 (both NEB) polymerases and a Techne TC-512 Thermocycler (Bibby Scientific Inc.) using the reaction mixture described in Table 2-5 and the reaction conditions described in Table 2-6.

Table 2-5: Reaction mixtures for SDM by PCR

Phusion™		Q5®	
10 µl	5x Phusion GC Buffer	10 µl	5x Q5 Reaction Buffer
2.5 µl	10 µM Forward Primer	2.5 µl	10 µM Forward Primer
2.5 µl	10 µM Reverse Primer	2.5 µl	10 µM Reverse Primer
1 µl	10 mM dNTPs	1 µl	10 mM dNTPs
0.5 µl	Phusion DNA Polymerase	0.5 µl	Q5 High-Fidelity DNA Polymerase
Up to 250 ng	Template DNA	Up to 1000 ng	Template DNA
Up to 33.5 µl	mQ water	Up to 33.5 µl	mQ water
50 µl	Final Volume	50 µl	Final Volume

Table 2-6: Reaction conditions for SDM by PCR

Step	Phusion™		Q5®	
Initial Denaturation	98°C	2 min	98°C	2 min
35 Cycles of: Denaturation	98°C	10 s	98°C	30 s
Annealing	45-72°C*	30 s	50-72°C*	30 s
Extension	72°C	30 sec per kb	72°C	1 min per kb
Final Extension	72°C	5-10 minutes	72°C	5-10 minutes
Hold	4°C		4°C	

*The T_m of the primer pair, as determined by NEB's T_m calculator was used for the Annealing temperature

All SDMs were initially attempted using Phusion™. After the PCR is complete, 1 µl of HF-DpnI was added to the reaction mixture and was incubated at 37°C for at least 2 hrs to remove the methylated template DNA before being analysed by agarose gel electrophoresis. The desired product were purified by gel extraction, as described in Section 2.2.3, then transformed directly into α-Select Bronze cells, as described in Section 2.2.4, and miniprepped, as described in Section 2.2.5.

If the SDM fails first time the PCR was repeated with Phusion™ but lowering the annealing temperature by 5°C. If it fails a second time, the PCR is attempted with Q5®, and if it still fails, then the primers were redesigned or Round-the-horn PCR was attempted, as described in section 2.2.7.

2.2.7 Round-the-horn (RTH) PCR

For more difficult SDMs, and for the removal of larger sections of DNA, Round-the-horn (RTH) PCR (OpenWetWare, 2013) was employed. Primers are purposely designed not to overlap, but are instead back to back in the sequence, with the desired mutation encoded in a single primer in the most 5' position, such that the end will not anneal to the wild-type sequence. Ideally T_ms of 60-72°C and GC contents of 50-60% were chosen.

This method of PCR results in blunt ended dsDNA (see Figure 2-1b) that requires ligating before transformation, so the primers need to be 5'-phosphorylated before the PCR. This was done using T4 Polynucleotide Kinase (PNK) (NEB). The reaction was set up as described in Table 2-7 below, incubated at 37°C for 30-60 min, then heat inactivated by heating to 65°C for 20 min.

Table 2-7: Reaction mixture for T4 Polynucleotide Kinase

	T4 Polynucleotide Kinase (PNK)
1 µl	10x T4 DNA Ligase buffer* (1 µM final concentration ATP)
1 µl	100 µM Forward Primer
1 µl	100 µM Reverse Primer
1 µl	PNK (1 unit**)
6 µl	mQ water
10 µl	Total Volume

*T4 PNK comes supplied with its own 10x buffer, which does not contain 10 µM ATP. NEB has stated that PNK functions to 100% efficiency in their T4 DNA Ligase buffer, which does contain ATP.
**1 unit is defined as the amount of enzyme catalysing the incorporation of 1 nmol of acid-insoluble [32P] in a total reaction volume of 50 µl in 30 minutes at 37°C in 1X T4 Polynucleotide Kinase Reaction Buffer with 66 µM [γ-32P] ATP (5 x 106 cpm/µmol) and 0.26 mM 5´-hydroxyl-terminated salmon sperm DNA.

PCR is then performed as described in Tables 2-5 and 2-6, except that 2.5 µl of the phosphorylated primer mixture, which contains 10 µM each of forward and reverse primer, and an additional 2.5 µl of mQ water were used in place of the normal 2.5 µl each of 10 µM forward and reverse primers.

After the PCR is complete, the mixture was DpnI digested, analysed and gel extracted as described in Section 2.2.6, then the gel extracted DNA was ligated using QuickStick Ligase as described in Table 2-4, before being transformed into α-Select Bronze cells, as described in Section 2.2.4, and miniprepped, as described in Section 2.2.5.

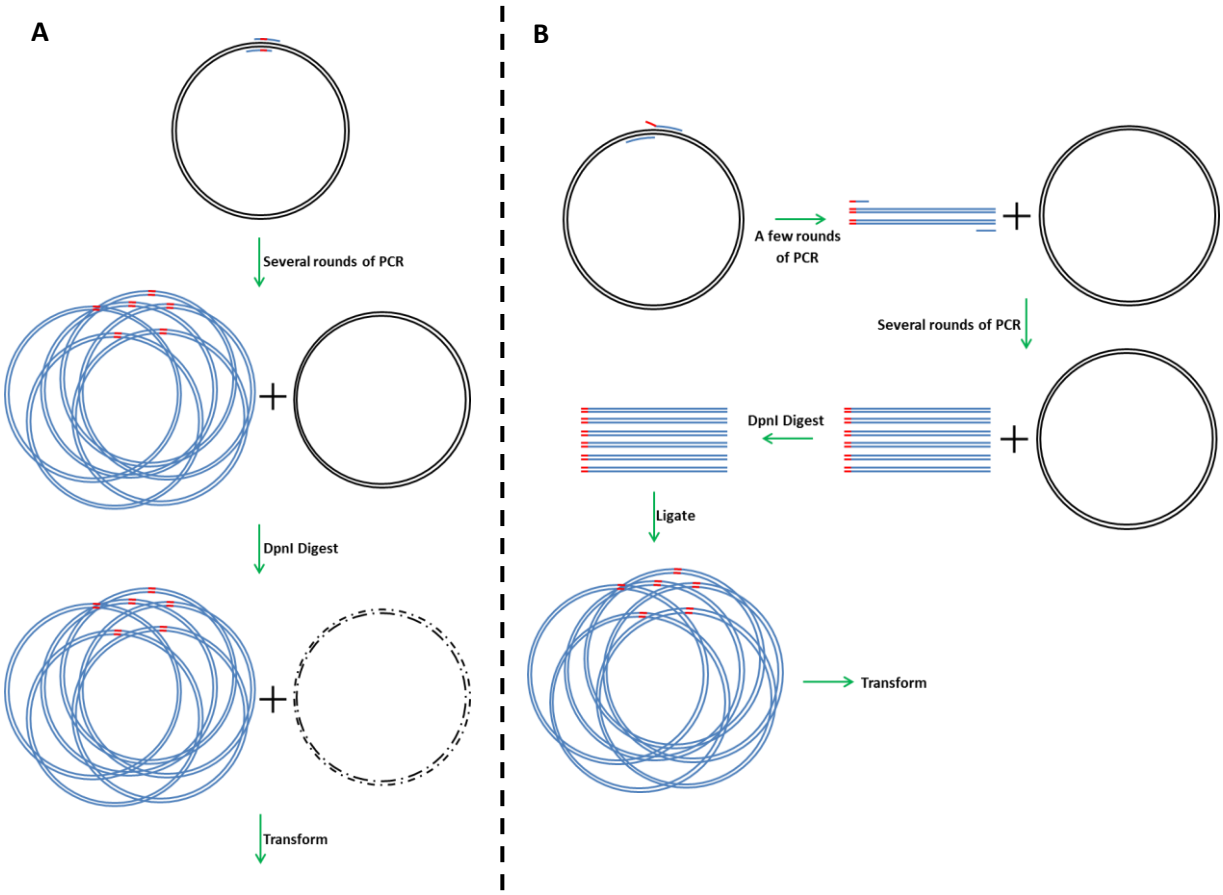


Figure 2-1 – Site-directed Mutagenesis (SDM): A comparison of the processes of "normal" SDM (A) and Round-the-horn (RTH) PCR (B). In normal SDM, the desired mutation (red) is incorporated into both PCR primers, whereas in RTH PCR, the mutation is present in only one primer and mutated region does not anneal to the template plasmid. (A) is better for small changes of one or a few base pairs, and (B) is better for larger mutations, deletions and insertions.

2.3 Protein Expression

For almost all recombinant protein expression, *E. coli* strain C41 competent cells were used. 50 µl aliquots were transformed with the plasmid of choice as described in Section 2.2.4. Colonies were collected from the agar plate and inoculated into fresh, sterilised 2xYT culture medium containing the appropriate antibiotic at 1 ml of stock solution per 1 l of medium and grown at 37°C in a shaking incubator until OD_{600} = 0.6. At this point they were induced with 0.5 mM IPTG and grown overnight (18 hrs.) at 25°C. Protein was either expressed on a small scale as a test of expression levels, using 5 ml of 2xYT in a 50 ml falcon tube, or on a large scale to create a stock for further experimentation, using 2 l or more of 2xYT in 0.5 l aliquots in a 2 l Erlenmeyer flask.

For large scale expressions, the cells were harvested by centrifuging at 8000 g for 8 minutes. The supernatant was decanted and the cell pellet was either immediately resuspended in the appropriate lysis buffer as described in Section 2.5.1, or collected and stored at -20°C for lysis at a later date.

2.4 Testing Protein Expression (Small Scale Preparation)

BugBuster® Master Mix (hence forth referred to simply as BugBuster) is a cell lysis agent that contains protein solubilising agents, including detergents, Benzoase® Nuclease and lysozyme designed to extract soluble proteins from cell pellets.

2 ml of culture from a 5 ml preparation was pelleted by centrifuging at 14000 g for 10 min at room temperature, and the supernatant discarded. The cell pellet was resuspended in 600 µl of BugBuster and incubated at room temperature for 20 min on a rotator. 30 µl of the lysate was removed as a sample of total cell protein to be analysed by SDS-PAGE later, before the lysate was pelleted by centrifuging at 14000 g for 10 min. 30 µl of the supernatant was removed as a sample of total soluble protein to be analysed by SDS-PAGE.

The insoluble fraction was then analysed by first discarding the supernatant and washing the pelleted debris twice by resuspending in 1 ml of 10% BugBuster in mQ water, repelleting by centrifuging at 14000 g for 10 min at room temperature, and discarding the supernatant. The remaining pellet was then resuspended in 300 µl of BugBuster and 30 µl was removed as a sample to be analysed.

2.5 Protein Purification (Large Scale Preparation)

2.5.1 Cell lysis

Fresh or frozen cell pellets were resuspended with 10 ml per 1 l of cell culture of the appropriate lysis buffer. The cell suspension was then lysed at 12000 psi using an EmusliFlex-C5 cell cracker (Avesin). A 10 µl sample of total cell lysate was taken for later analysis by SDS-PAGE, as described in Section 2.5.4. The lysate was then centrifuged at 35000 g for 35 min at 4°C, and the supernatant containing the soluble proteins was collected. A 10 µl sample of total soluble protein was taken for SDS-PAGE analysis.

2.5.2 Batch Purification

For proteins containing a thrombin-cleavable His-tag, nickel-NTA-agarose beads (Generon) (1 ml + 1 ml per litre of culture) were washed with 50 ml of the appropriate wash buffer, to remove storage solution. The washed beads and the supernatant from Section 2.5.1 were combined and incubated for 1-2 hrs at 4°C on a nutator, rotating at 8 rpm. The mixture was then pelleted by centrifugation at 1000 g for 5 min at 4°C, the supernatant, containing the unbound proteins, was collected and retained for SDS-PAGE analysis. The Ni-NTA beads with protein bound were then washed three

times by resuspending in 45 ml of Wash buffer, and then pelleted by centrifugation at 1000 g for 5 min at 4°C. Each time the supernatant was collected and retained for SDS-PAGE analysis.

The Ni-NTA beads were then resuspended in 5 ml of wash buffer and a 30 µl sample of the matrix before thrombin cleavage was taken for SDS-PAGE analysis. 25 units per 1 l of culture of thrombin from bovine plasma were added and then incubated for overnight at 4°C, rotating at 8 rpm to facilitate thrombin cleavage of the protein from the His-tag.

Afterwards 1 ml of 1 M imidazole-HCl pH 7.4 was added to raise the imidazole concentration to approximately 200 mM and the suspension was mixed for a further 1 min before being pelleted by centrifugation at 1000 g for 5 min at 4°C. The supernatant was collected as the first elution. Two more elutions were obtained by resuspending the beads in 5 ml of the appropriate Elution buffer, mixing for 1 min, then pelleting by centrifugation at 1000 g for 5 min at 4°C, and collecting the supernatant.

For proteins containing a thrombin-cleavable GST-tag, a similar protocol was used with the following exceptions: glutathione sepharose beads (Generon) were used in place of Ni-NTA-agarose beads; the beads with proteins bound were washed three times with High Salt Wash buffer and three times with Low Salt Wash buffer for a total of six washes; and the elutions were done using just Low Salt Wash Buffer

2.5.3 Further Purification using Column Chromatography

Soluble protein aggregates and DNA were removed using column chromatography on an Äkta Pure or an Äkta Purifier (GE Healthcare). For TCF7L2 constructs this was done by size-exclusion chromatography (SEC). A HiLoad™ 26/60 Superdex™ 75 column (G75) (GE Healthcare) was pre equilibrated with G75 buffer. The three elutions were combined, filtered using a Millex Express PES 0.22 µm filter (Merck) and loaded onto the G75, and eluted with 1.2 column volumes (CV) of G75 buffer. The fractions corresponding to A_{215} peaks were analysed by SDS-PAGE and the fractions containing >95% of the protein of interest were combined, then divided into aliquots, flash frozen with liquid nitrogen and stored at -80°C until use.

For the Weis β-catenin construct, additional purification was done by anion exchange. A Mono-Q HR 10/10 anion exchange column (MonoQ) (GE Healthcare) was pre-equilibrated with 3 CV of Low Salt MonoQ buffer. The three elutions were combined, diluted up to 45 ml, 0.22 µm filtered as above, loaded onto the MonoQ, and eluted using a linear gradient of High Salt MonoQ buffer over 15 CV. The fractions corresponding to A_{280} peaks were analysed by SDS-PAGE and the fractions containing

>95% of the protein of interest were combined. Purified β-catenin was snap frozen with dry ice and lyophilised under vacuum before being stored at -80°C until use.

2.5.4 Protein Analysis by SDS-PAGE

Samples from the output fractions from the Äkta were analysed in order to determine which contain the protein of interest, along with those taken throughout the purification process, in order to troubleshoot any potential problems, such as low yield. The samples were prepared as described in Table 2-8 below, however some samples contain higher concentrations of protein than others and so smaller volumes of these samples were used (see Table 2-9).

The SDS-PAGE samples were heated 80°C for 10 min then 18 µl loaded into the wells of the gel. NuPAGE® Novex 4-12% Bis-Tris gels (Novagen) were used for LRRK2 constructs, RunBlue Bis-Tris 4-12% gels (Expedeon) were used for TCF7L2 constructs and RunBlue SDS 4-20% gels (Expedeon) were used for β-catenin. Bis-Tris gels were run using their respective 1x MES-SDS at 200 V and 120 mA per gel for 45 min, the 4-20% SDS gels were run using 1x Teo-Tricine at 180 V and 110 mA per gel for 60 min. Proteins were visualised with Coomassie Brilliant Blue G-250 as described Lawrence and Besir (2009) and the gels photographed under white light using an Alpha Imager.

Table 2-8: SDS-PAGE sample preparation

SDS-PAGE Samples	
5 µl	4x SDS Loading Buffer
1-10 µl	Sample
1 µl	1 M DTT
Up to 13 µl	mQ water
20 µl	Total Volume

Table 2-9: Amount of sample used in SDS-PAGE

Sample	Amount used
Total cell lysate, total soluble protein, unbound protein	1 µl
Wash 1/high salt wash 1, matrix before and after thrombin	5 µl
All other washes, all elutions, all Äkta fractions, all BugBuster fractions	10 µl

2.5.5 Estimation of Protein Concentration

An estimation of protein concentration was obtained by measuring the absorbance at 280 nm (A_{280}) using a NanoDrop 2000 UV-Vis spectrophotometer. The concentration of a substance can be calculated using the Beers Law which can be rearranged thus:

$$C = \frac{A_x}{\varepsilon_x l} \qquad [2\text{-}1]$$

Where C is the concentration in M, A_x is the absorbance value at x nm, ε_x is the extinction coefficient of that substance at x nm in M^{-1} cm^{-1}, and l is the pathlength in cm.

The residues of tyrosine and tryptophan, and to a lesser extent phenylalanine and Cystine, the diamino acid formed by the formation of a disulfide bridge between two cysteine residues, absorb strongly at 280 nm. An estimation of ε_{280} for each of my constructs was obtained using ProtParam (Gasteiger *et al.*, 2005) which uses the Edelhoch Method (Edelhoch, 1967):

$$\varepsilon_{280}(Protein) = Numb(Tyr) \times \varepsilon_{280}(Tyr) + Numb(Trp) \times \varepsilon_{280}(Trp) + Numb(Cystine) \times \varepsilon_{280}(Cystine) \quad [2\text{-}2]$$

Where ε_{280}(Tyr) = 1490, ε_{280}(Trp) = 5500, and ε_{280}(Cystine) = 125. (Pace *et al.*, 1995)

Absorbance values for the NanoDrop are correct for a pathlength of 1 cm and the estimations of the ε_{280}s of my constructs are listed in Table 2-10 below.

Table 2-10: Extinction coefficients of protein constructs used in this Thesis

Protein	ε_{280} / M^{-1} cm^{-1}
LRRK2 ARM Construct	43400
LRRK2 ANK Construct	14480
LRRK2 LRR Construct	26065
LRRK2 WD40 Construct	44055
Weis β-catenin	32025
TCF7L2	-

The TCF7L2 constructs only contains a single phenylalanine which does not produces a strong enough absorbance signal to allow concentration to be determined by this method. The concentration of TCF7L2 stocks were estimated by Bicinchoninic Acid (BCA) assay, using a Pierce™ BCA protein Assay Kit (ThermoFisher) and determined more accurately by ion-exchange-ninhydrin analysis performed by Dr. Peter Sharratt at the University of Cambridge PNAC Facility.

2.5.6 Confirmation of Protein Identity

Purified proteins were analysed by ion-exchange-ninhydrin analysis as described in Section 2.5.5 to determine amino acid composition and protein concentration and by matrix-assisted laser desorption/ionisation time of flight (MALDI TOF) mass spectroscopy, performed by Dr. Len Packman also at the PNAC Facility. Corroboration between these two analytical techniques was taken as evidence that the isolated protein was the protein of interest.

2.6 Fluorescent Labelling

2.6.1 Labelling Lysine Residues with Fluorescein Isothiocyanate (FITC)

The isothiocyanate moiety reacts with amines and also sulfhydryl groups by a simple addition reaction, reacting faster at the less crowded ε-amines of lysine residues than the N-terminal amine group.

A 500 µl aliquot of TCF7L2 (1-54) in 1x PBS was buffer exchanged using a Centri Pure P5 desalting column (Generon) into a buffer containing 10 mM phosphate between pH 9 and 10, and 1 M NaCl. This is to ensure that the amine moieties were deprotonated. 10 mM FITC in 100% DMSO was added to a 5 times molar excess and incubated at 37°C in an opaque, black microfuge tube in the dark for at least 3 hrs.

Excess unreacted FITC was removed by buffer exchange into 1xPBS using a PD MidiTrap G-10 column (GE Healthcare). The collected fractions with analysed by SDS-PAGE, visualised and photographed under UV light before Coomassie staining, then visualised and photographed again under visible light after Coomassie staining (see Figure 2-2). Fractions that contained protein of the correct molecular weight, which also fluoresced under UV light were collected and pooled. Excess FITC ran at the same molecular weight as the dye front. Final concentration of protein was determined by BCA assay and final label concentration was determined by Nanodrop, using the extinction coefficient ε_{519} = 68000. Labelling efficiency was determined by comparison of label and protein concentrations.

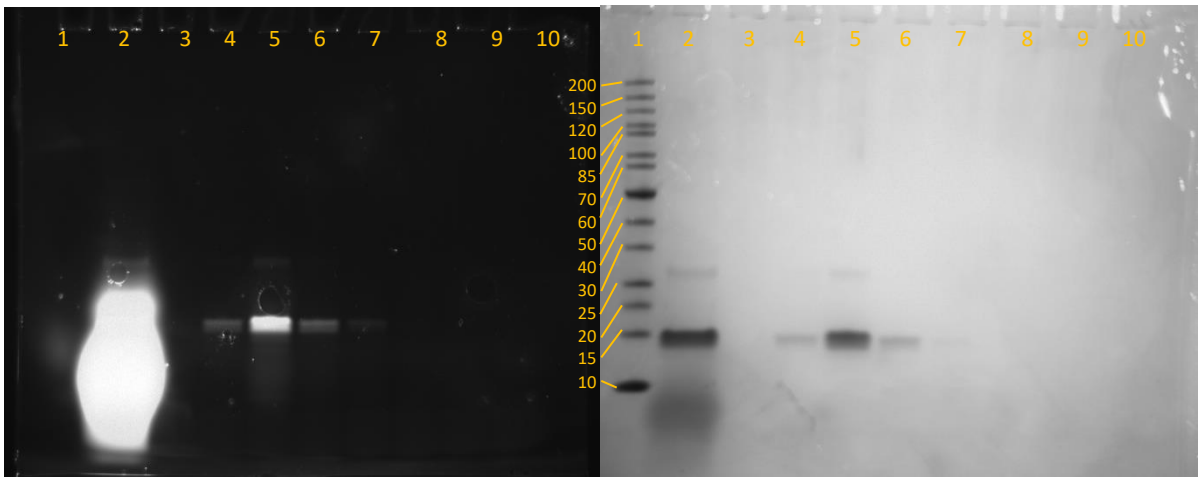


Figure 2-2 – SDS-PAGE analysis of FITC labelled TCF7L2 (1-54). Visualised with UV-light before Coomassie Blue staining (left) and white light after Coomassie Blue staining (right). Lane 1 – ladder; 2 – labelling reaction mixture; 3-10 – fractions 1-8.

2.6.2 Labelling Cysteine Residues with Fluorescein Maleimide

The maleimide functional group reacts with sulfhydryl groups via nucleophilic conjugation addition and so can be used to label proteins at cysteine residues. A 500 µl aliquot of TCF7L2 (1-54) construct containing a cysteine residue in 1x PBS buffer with 1 mM DTT was buffer exchanged into fresh 1x PBS without DTT, as described in Section 2.6.1. Fluorescein maleimide also reacts with the sulfhydryl groups present in DTT, which is in vast excess of the protein and so needs to be removed before labelling. 20 mM fluorescein maleimide in 100% DMSO was added to a 5 times molar excess and incubated at 37°C in an opaque, black microfuge tube in the dark for at least 1 hr.

Excess fluorescein maleimide was removed using a method developed by Vivès and Lebleu (2003). 10x volume of acetone, pre-chilled to -20°C, is added to the labelling reaction mixture, mixed by vortex, and incubated at -20°C for 30 min. Fluorescein, and its derivatives, are very soluble in acetone, whereas peptides and proteins are insoluble in acetone. The addition of acetone to 90+% causes the protein to precipitate out of solution, while the unreacted fluorescein maleimide remains in solution. The reaction mixture was centrifuged at 15000 g at 4°C for 10 min to collect the precipitated, labelled, protein and the supernatant removed and the pellet was resuspended in 1x PBS. This acetone wash process is repeated twice more but the precipitated protein pellet was not resuspended on the final wash. The pellet is left at room temperature, in the dark, and open to the air for 16 hrs to ensure that any residual acetone has evaporated away, after which the desiccated labelled-TCF7L2 can be stored at 4 °C for up to 5 days.

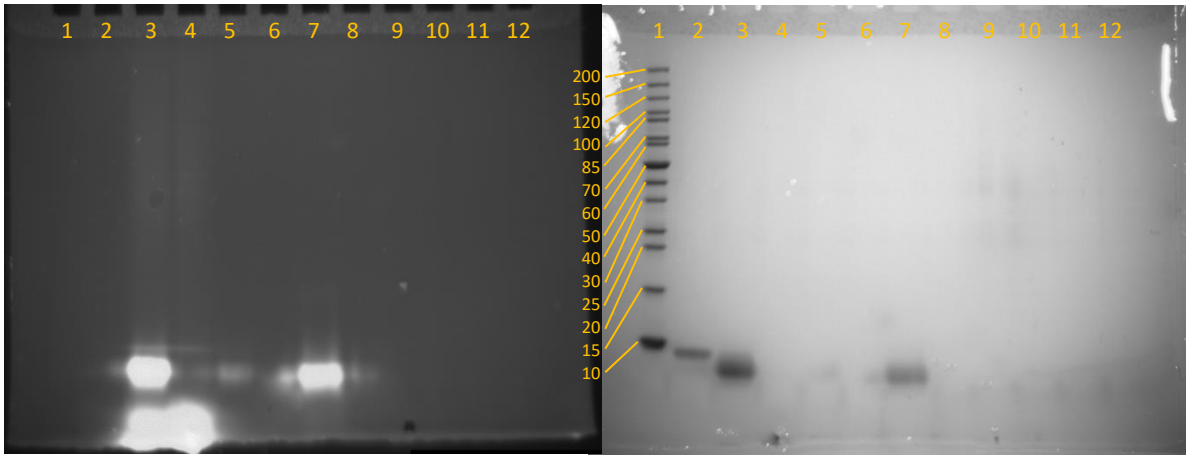


Figure 2-3 – SDS-PAGE analysis of fluorescein maleimide labelled TCF7L2 S31C (1-54). Visualised with UV-light before Coomassie Blue staining (left) and white light after Coomassie Blue staining (right). Lane 1 – Ladder; 2 – Unlabelled TCF7L2 S31C stock; 3 – labelling reaction mixture; 4-6 – supernatant from acetone washes 1-3; 7 – resuspended TCF7L2 after 3 washes; 8-12 – empty lanes.

2.7 Isothermal Titration Calorimetry (ITC)

ITC measurements were performed using VP-ITC Calorimeter (Malvern, formally Microcal). Both β-catenin and TCF7L2 construct after labelling are stored as lyophilised powders. To ensure matched buffers, lyophilised β-catenin was resuspended in mQ water and the buffer exchanged into fresh 50 mM Tris pH 8.9, 100 mM NaCl 2 mM DTT using Centri Pure P5 desalting column and the TCF7L2 was resuspended in the same buffer. Unlabelled TCF7L2 was washed with acetone three times, as described in Section 2.6.2 to produce the lyophilised pellet. β-catenin was used at a concentration of 2-4 μM and TCF7L2 was used at 6-10x higher concentration (18-40 μM). Blank titrations of TCF7L2 into water and water into β-catenin were performed to demonstrate that the reaction observed was a binding interaction. All experiments were performed at 30°C.

Data was fitted using Origin software package (Malvern, formally Microcal) using the one site binding equation to calculate the equilibrium association constant, K_a, enthalpy change, ΔH, entropy change, ΔS, and stoichiometry, n, of binding.

2.8 Kinetic Studies by Stopped-Flow Spectroscopy

Stopped-flow spectroscopy was performed using a SX-19 Stopped flow fluorimeter (Applied Photophysics) in fluorescence mode. Excitation and emission wavelengths were set to 495 nm and 519 nm respectively and a 515 nm long-pass filter was used to improve the signal to noise ratio. All proteins were prepared in fresh 1x PBS buffer with 1 mM DTT and the experiments performed at 15°C, except when otherwise stated.

For association experiments, a fixed concentration of labelled TCF7L2 construct was rapidly mixed, in a 1:1 ratio, with varying concentrations unlabelled β-catenin and the change in fluorescence intensity was recorded. The β-catenin concentrations were chosen to be at least ten times higher than that of the TCF7L2 to ensure a pseudo-first order association reaction. A final concentration 10 nM TCF7L2 was used and the final concentration of β-catenin was varied between 100 nM and 1000 nM.

For dissociation experiments, a complex of labelled TCF7L2 and β-catenin was pre-formed by mixing to two proteins in a 1:1 molar ratio and incubating them at 15°C for 1 hr in the dark to reach equilibrium. The pre-formed complex was then mixed in a 1:1 ratio with 50 times molar excess unlabelled WT TCF7L2 and the change in fluorescence intensity was measured. The final concentration of complex used was 200 nM with a final competitor concentration was 10 µM.

A total of five traces for each concentration of β-catenin were collected, and a minimum of three traces, usually all five, were averaged. The average curve was plotted using GraphPad Prism 6 (GraphPad Software, Ltd), and fitted with either a one- or two-phase exponential equation:

$$y(t) = Ae^{-kt} + bt + c \quad [2\text{-}3]$$

$$y(t) = A_1e^{-k_1t} + A_2e^{-k_2t} + bt + c \quad [2\text{-}4]$$

Where y(t) is the fluorescence intensity at time, t; A_n and k_n are the amplitude and rate constant of phase n, respectively; b is correction factor for linear drift as a result of photo-bleaching (only relevant for very long recordings) and c is the fluorescence intensity when the trace has reached a plateau. In all cases the first 2 ms of each trace was excluded from the fit to reduce error brought about by incomplete mixing of the two binding partners.

For association experiments, the observed rate constant, k_{obs}, was then plotted against the concentration of β-catenin and k_{on} calculated from the linear fit:

$$k_{obs} = k_{on}[\beta\text{catenin}] + c \quad [2\text{-}5]$$

2.9 Kinetic Studies by Fluorescence Spectroscopy

After it was found that the stopped-flow was unsuitable for recording over sufficiently long time scales to measure complex dissociation, time-dependent fluorescence spectroscopy was performed using a LS55 Luminescence Spectrometer (Perkin-Elmer) using a cell of pathlength 10 mm. Excitation and emission wavelengths were set to 495 nm and 519 nm respectively, shutter width was 5 nm and excitation voltage was 650 V. All proteins were prepared in fresh 1x PBS buffer with 1 mM DTT and the experiments performed at 15°C, except when otherwise stated.

For dissociation experiments, a complex of labelled TCF7L2 and β-catenin was pre-formed by mixing to two proteins in a 1:1 molar ratio and incubating them at 15°C for 1 hr in the dark to reach equilibrium. The pre-formed complex was then mixed in a 1:1 ratio with 50 times molar excess unlabelled WT TCF7L2 and the change in fluorescence intensity was measured. Individual traces were fitted with Equations [2-3] and [2-4].

2.10 Equilibrium Studies by Fluorescence Anisotropy

Fluorescence spectroscopy was performed using a LS55 Luminescence Spectrometer (Perkin-Elmer) using a cell of pathlength 10 mm. Excitation and emission wavelengths were set to 495 nm and 519 nm respectively, shutter width was 10 nm and excitation voltage was 750 V. All proteins were prepared in fresh 1x PBS buffer with 1 mM DTT and the experiments performed at 15°C, except when otherwise stated.

Titrations of β-catenin in 42 steps were performed using a Microlab 500C series dispenser (Hamilton Company). These titrations were divided into 20 steps of 1 µl, followed by 10 steps of 2 µl, followed by 12 steps of 5 µl, for a total of 100 µl added to 1 ml of TCF7L2 construct in a cuvette. The solution in the cuvette was stirred for 30 s then incubated for a further 60 s before the fluorescence intensity and polarisations were recorded.

Anisotropy is measured as:

$$r = \frac{(I_{VV} - GI_{VH})}{(I_{VV} - 2GI_{VH})} \quad [2\text{-}6]$$

Where I_{VV} is the fluorescence intensity with both polarisers aligned vertically, I_{VH} is the fluorescence intensity with the excitation polariser aligned vertically and the emission polariser aligned horizontally. G is the grating factor which is measured as:

$$G = \frac{I_{HV}}{I_{HH}} \quad [2\text{-}7]$$

Where I_{HV} is the fluorescence intensity with the excitation polariser aligned horizontally and the emission polariser aligned vertically and I_{HH} is the fluorescence intensity with both polarisers aligned horizontally. The grating factor is a correction factor to allow for differences in instrument polarisation. For these experiments, 1.09 ≤ G ≤ 1.12, and titrations did not commence until G measurements were stable to 2 d.p.s.

Fluorescence measurements need to be corrected for dilution. All data was plotted in GraphPad Prism 6 and was fitted to the following equations:

Polarisation, accounting for ligand depletion –

$$r = r_{min} + \frac{(r_{max} - r_{min}) \times (b - \sqrt{b^2 - 4c})}{2a} \quad [2\text{-}8]$$

Where r is the anisotropy reading, $a = [TCF7L2]$, $b = [TCF7L2] + [\beta\text{catenin}] + K_d$, $c = [TCF7L2] \times [\beta\text{catenin}]$, r_{min} is the free ligand polarisation, and r_{max} is the saturation polarisation, for constant [TCF7L2].

Hills Equation -

$$r = r_{min} + \frac{(r_{max} - r_{min}) \times [\beta\text{catenin}]^H}{({K_d}^H + [\beta\text{catenin}]^H)} \quad [2\text{-}9]$$

Where r is the anisotropy reading, r_{min} is the free ligand polarisation and r_{max} is the saturation polarisation, and H is the Hill's coefficient.

Mono-Wyman-Changeux (MWC) model -

$$r = r_{min} + (r_{max} - r_{min}) \times \frac{[\beta\text{catenin}]K^R(1+[\beta\text{catenin}]K^R)^{n-1} + Lc[\beta\text{catenin}]K^R(1+c[\beta\text{catenin}]\times K^R)^{n-1}}{(1+[\beta\text{catenin}]K^R)^n + L(1+c[\beta\text{catenin}]K^R)^n} \quad [2\text{-}10]$$

The MWC model (Monod, Wyman and Changeux, 1965; Stefan, Edelstein and Le Novère, 2009) is another model which accounts for cooperativity in multiple ligand binding. The MWC model assumes that there are n identical ligand-binding sites, where each undergo concerted allosteric transitions. These binding sites are assumed to adopt two conformations, a "tense", low ligand-affinity conformation, "*T*", and a "relaxed", high ligand-affinity conformation, "*R*". *L* in Equation 2-10

is a constant defined as the ratio of binding sites in the "tense" conformation to those in the "relaxed" conformation in the absence of ligand: $L = [T_0]/[R_0]$. K^R is the association constant for the "relaxed" state and the constant c in Equation 2-10 is the ratio of association constants for "tense" and "relaxed" states $c = K^T/K^R$.

2.11 Errors

The error in the sum a + b = c, γ

$$\gamma = \sqrt{\alpha^2 + \beta^2} \quad [2\text{-}11]$$

The error in the product a x b = c, γ

$$\gamma/c = \sqrt{(\alpha/a)^2 + (\beta/b)^2} \quad [2\text{-}12]$$

The error in the exponential exp(a) = c, γ

$$\gamma = a \times \alpha \quad [2\text{-}13]$$

The error in the logarithm ln(a) = c, γ

$$\gamma = \alpha/a \quad [2\text{-}14]$$

For a set of *n* measurements ($a_i \pm \alpha_i$), the weighted mean, <a> is given by:

$$< a >= \sum_{i=1}^{n} a_i/{\alpha_i}^2 \Big/ \sum_{i=1}^{n} {\alpha_i}^{-2} \quad [2\text{-}15]$$

And the error in the weighted mean, <a>, <α>

$$< \alpha >= \sqrt{\sum_{i=1}^{n} ((a_i - < a >)/\alpha_i)^2 \Big/ (n-1) \sum_{i=1}^{n} {\alpha_i}^{-2}} \quad [2\text{-}16]$$

Chapter 3 – LRRK2: From Model to Molecule

The original intended subject of my studies was the protein Leucine-Rich Repeat Kinase 2, hereafter referred to as LRRK2. This very large, 2527 residue protein has been strongly linked to familial Parkinson's disease. It is believed to be comprised of seven domains: A MAPKKK-like, serine/threonine kinase domain; a Roc (Ras of complex proteins) domain encoding a GTPase; a COR (C-terminal of Roc) domain forming a catalytic core; and four protein-protein interaction domains predicted to be ARM repeat, ANK repeat, LRR and WD40 repeat domains (Mills *et al.*, 2012). At the time of writing, the structure of this protein is largely unsolved, with only the isolated COR-domain being solved (PDB ID 2ZEJ) (Deng *et al.*, 2008).

This project originated from the paper, ***Analysis of LRRK2 Accessory Repeat Domains: Prediction of Repeat Length, Number and Sites of Parkinson's Disease Mutations*** by Mills *et al.* (2012). In this paper Mills *et al.* used homology modelling to predict the structure of the four repeat domains. The purpose of this project was to biophysically and structurally characterise these domains, both in isolation and in the context of the whole protein.

LRRK2 has been shown to interact with a large number of different protein partners (Kang and Marto, 2017). So after these domains have been characterised the goal was then to determine where and how a number of these partner interact. This can then serve as basis for the design of small molecule inhibitors for those interactions.

In this Chapter I describe my efforts to design, express and purify proteins constructs corresponding to the four tandem repeat domains described in Mills *et al.* (2012), first in isolation, then by co-expressing the individual constructs.

3.1 Designing LRRK2 Constructs

The full-length protein sequence was taken from UniProt's Protein Knowledgebase, ID Q5S007, and compared to the proposed repeat-containing regions described in Mills *et al.*, (2012) (See Appendix A, Section A.1). The main consideration, when designing the initial constructs, was that the predicted locations of the repeats may be inaccurate and that repeat proteins tend to have accessory loops N- and C-terminal of the repeating region to aid overall stability, so as many as was reasonable amino acids were included either side of the repeating region. The other main consideration during the initial design was that I wanted to study the domains in isolation from each other, so the constructs were designed such that their sequences did not overlap. The nucleotide base sequence used was taken from NCBI's Gene Bank, file 120892.

The ARM repeat domain is the largest and most N-terminal of the predicted domains, with fourteen proposed repeats of 42 aa (residues 49-657 aa). The construct designed to represent this domain comprises residues 1-661 aa. The ANK repeat domain is the next domain predicted in the sequence and is also the smallest with seven repeats of 33 aa (residues 676-902 aa). The construct designed to represent this domain comprises residues 668-928 aa. The LRR repeat domain is the last of the predicted N-terminal domains with fourteen repeats of 24 aa (residues 984-1320 aa). The construct designed to represent this domain comprises residues 979-1328 aa. Finally the WD40 repeat domain is predicted to be at the C-terminus, after the catalytic domains, and is comprised of 7 repeats of 40 aa (residues 2164-2515 aa). The construct designed to represent this domain comprises residues 2159-2527 (see Appendix A, Sections A.1.1-4).

All four constructs were given a 5' (N-terminal) methionine codon and a 3' (C-terminal) stop codon where appropriate. They were also given BamHI and HindIII cloning sites as the 5' and 3' end respectively.

An estimate of the biophysical properties of the four artificial constructs was obtained using ExPASy ProtParam (Gasteiger *et al.*, 2005). The key properties of interest are: the molecular weight, so the construct proteins could be identified by SDS-PAGE; the extinction coefficient at 280 nm (ε_{280}), so the concentration of protein could be obtained by spectrophotometry; and the theoretical isoelectronic point (pI) as this represents the pH at which there is no net charge on the protein and therefore buffers should have a pH at least 1 unit above or below this value. The pIs of my four constructs were 5.48, 4.99, 6.50 and 8.93 for ARM, ANK, LRR and WD40 respectively, so a pH of 7.4 was chosen for buffers. While not ideal for the LRR construct it did at least serve as a starting point.

It also so happened that all of my constructs naturally contain at least one tryptophan residue, so it may be possible to measure biophysical properties through intrinsic fluorescence (see Table 2-10 in Section 2.5.5), providing they are in a solvent-responsive microenvironment. They also all contain cysteine and lysine residues for labelling with dyes if intrinsic fluorescence is not suitable. The codon usage was optimised for *E. coli* expression whilst conserving BamHI and HindIII cloning sites using GeneArt (Life Technologies).

3.2 Initial Expression Tests Showed Proteins were Present but Insoluble

Once all of my constructs had been subcloned into expression vectors, a small-scale expression was performed as described in Section 2.3, and the proteins extracted with BugBuster according to Section 2.4 except that a sample of total cell protein was also taken from the pre induced cells. Each of the individual constructs expressed but none of them showed a significant degree of soluble

expression (see Figure 3-1). Three of the constructs, His-tagged WD40, GST-tagged ANK and GST-tagged LRR, showed a significant amount of insoluble expression (lane 1 of (B), (C) and (E) respectively).

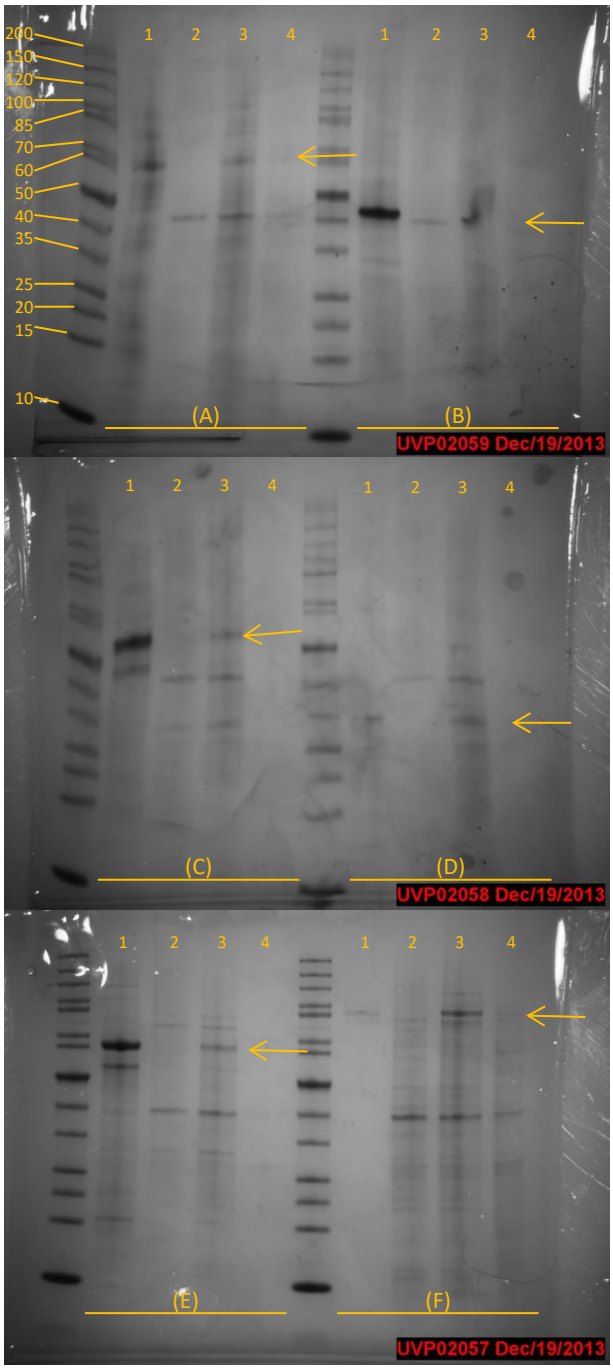


Figure 3-1 – Initial expression test of four LRRK2 constructs in different expression vectors. Visualised by Coomassie blue staining (A) WD40 construct in pRSET-GST vector ($M_w \approx 67.5$ kDa); (B) WD40 construct in pRSET vector ($M_w \approx 41.5$ kDa); (C) ANK construct in pRSET-GST vector ($M_w \approx 52.7$ kDa); (D) ANK construct in pRSET vector ($M_w \approx 26.7$ kDa, but runs higher); (E) LRR construct in pRSET-GST vector ($M_w \approx 66.1$ kDa); (F) ARM construct in pRSET-GST vector ($M_w \approx 100.5$ kDa). The expected position is denoted by ←. Lanes marked 1 are the insoluble protein fraction; 2 are the soluble protein fraction; 3 are the total cell protein after induction with IPTG; and 4 are the total cell protein before induction. The protein ladder was the same for all gels.

Hexa-His tagged WD40 ((B) in Figure 3-1) appeared to show the highest expression but it was difficult to tell the relative level of soluble expression due to the ubiquitous band at ≈ 40 kDa, and so a larger-scale expression was attempted, performed as described in Section 2.3. Two different methods were used in an attempt to isolate and purify the soluble fraction. First an attempt was made using SEC as described in Section 2.5.3 using a HiLoad™ 26/60 Superdex™ 75 column, and the second was made using a Mono S 5/50 GL (GE Life Sciences) cation exchange column in the place of the MonoQ. Neither attempt was successful, as there was too little soluble protein to be isolated (see Figure 3-2). Whilst washing the Mono S column with 20% ethanol afterward, a large peak was observed within one column volume that could have corresponded to protein.

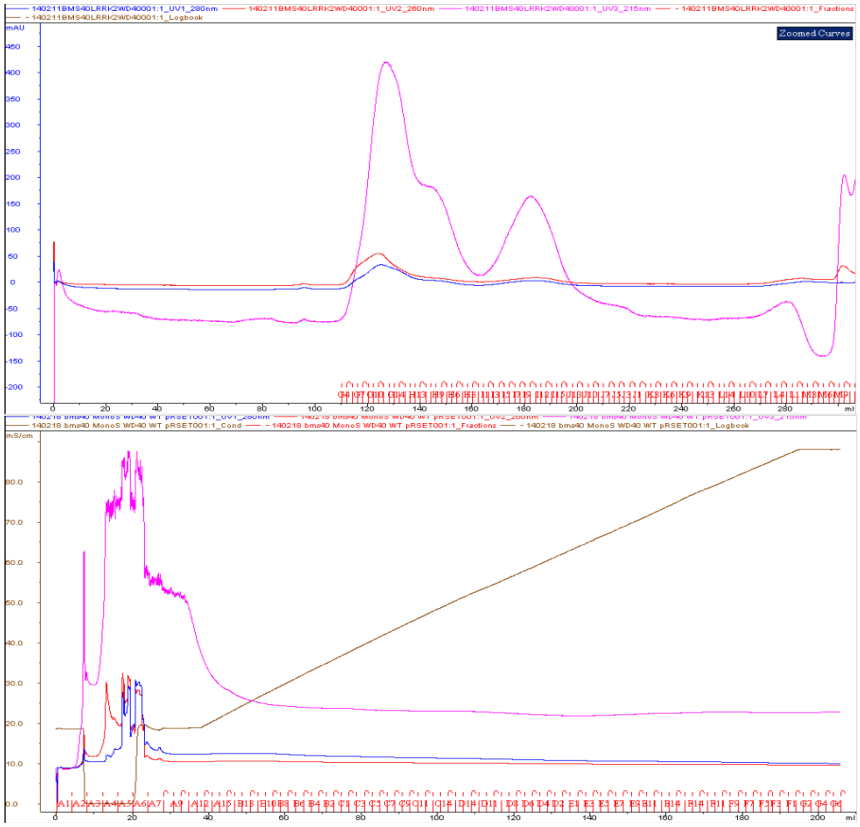


Figure 3-2 – Analysis of larger scale expressions. The top trace is the elution profile from the first attempt to purify WD40 in pRSET. The graph shows how the absorption at 215 nm (purple - peptide bond), 260 nm (red - nucleic acids) and 280 nm (blue - protein fluorescent residues) changed with elution volume. The second trace is the elution profile from the attempt to purify WD40 in pRSET by Mono S.

A third large scale purification attempt was made, this time using the GST-fused ANK construct ((C) in Figure 3-1) as described in Sections 2.3 and 2.5.2. The samples, including the insoluble fraction were analysed by SDS-PAGE before Äkta purification was attempted. The gel revealed that there was no ANK construct protein present in any of the elutions and only a very small amount was present on the glutathione sepharose matrix after thrombin cleavage. Almost all of the protein was in the insoluble fraction (see Figure 3-3, lane 3).

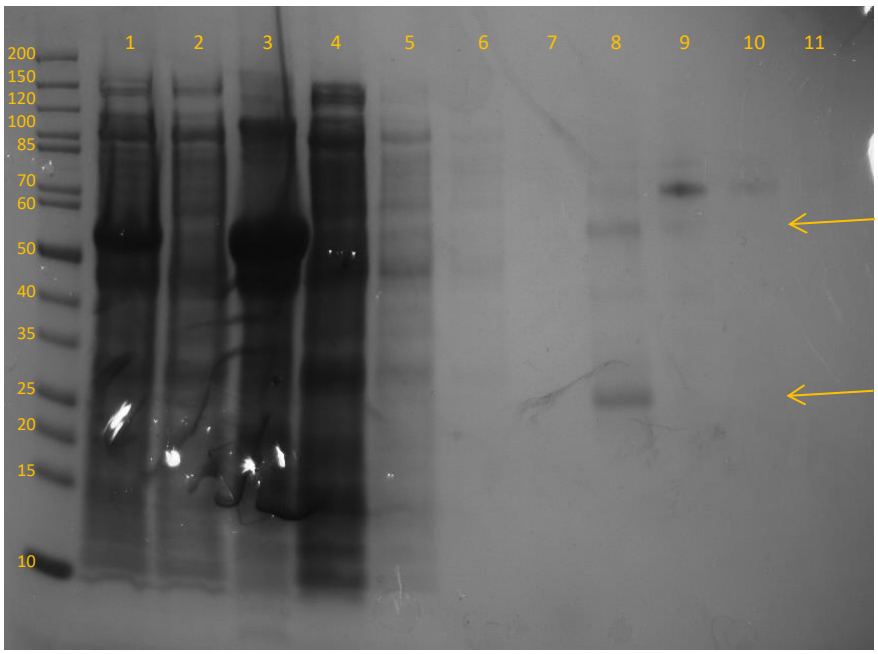


Figure 3-3 – SDS-PAGE analysis of the attempted purification of GST-fused ANK. Lane 1 - Total cell protein; 2 - Total soluble protein; 3 - Total insoluble protein; 4 - Unbound protein; 5-7 - Matrix washes 1-3; 8 - Matrix after thrombin incubation; 9-11 - Elutions 1-3. The top ← indicates the expected position of the GST-ANK fusion protein and the bottom indicates that of just the ANK protein.

3.3 Screening Lysis Additives

A recent paper by Leibly *et al*. (2012) has shown that many recombinant proteins only become insoluble during the lysis stage of recombinant protein purification, as chaperones and other cellular stabilising agents are removed or destroyed, and that a number of compounds could be added to the lysis buffer to successfully obtain soluble protein from previously insoluble expression. A list of 144 additives was compiled, from which they obtained a short list of sixteen that were successful in rescuing multiple recombinant proteins (see Table 3-1).

These sixteen additives were added to BugBuster, to the concentration described in Table 3-1 and a 1 l expression of ANK in pRSET was prepared as described in Section 2.3. The cell pellet was then washed with 500 ml 1x PBS, repelleted, and then resuspended in 500 ml of 1x PBS. 2 ml aliquots were removed and the protein extracted as described in Section 2.4 using the BugBuster containing the various additives and the fractions were analysed by SDS-PAGE. None of the additives had the desired effect of solubilising the ANK construct, although it is interesting to note that sodium selenite resulted in the band predicted to be the ANK construct running lower than it had for the other additives (see Figure 3-4). Nothing in the literature make a reference to this observed effect.

Table 3-1: Short list of additives used by Leibly *et al.*.

Compound	Concentration	Compound Classification
Trehalose	0.75 M	Naturally occurring sugars
Mannitol	0.5 M	
Xylitol	1 M	
Glycine betaine	1 M	Zwitterionic osmolytes
TMAO	1 M	
L-proline	0.5 M	
L-arginine	0.375 M	
NDSB 195	1 M	Non-naturally occurring zwitterions
NDSB 201	1 M	
Trisodium citrate	0.1 M	Naturally occurring salts
Dipotassium hydrogen phosphate	0.1 M	
Sodium selenite	0.01 M	Inorganic salt
Copper (II) chloride	0.01 M	Metal ion
CTAB	0.5% (w/v)	Detergents
SDS	0.1% (v/v)	
Triton X-100	0.5% (v/v)	

3.4 Troubleshooting the Lack of Solubility

There are a number of reasons that could explain why the designed constructs are insoluble whereas the parent protein is soluble. The first, in this case, is the fact that repeat proteins tend to very dependent on the stability of the terminal repeats. The internal repeats tend to be enriched with hydrophobic residues as large regions of the repeats' surfaces tend to be packed against neighbouring repeats, and are reliant on these interactions with their neighbours for stability. Whereas the terminal repeats tend to be slightly more stable as they have only have one neighbour which can contribute to stability. So when the terminal repeat is destabilised to the point where it can no longer fold, it leads to a cascade of unfolding as the internal repeats lose stability provided by their neighbour as they successively unfold. Generally repeat proteins have extended terminal loops that fold across the first/last repeat giving increased stability. For my constructs, this could be rectified with additional subcloning to extend the loops and was the first solution attempted, however for some of the constructs it was not possible to provide an extended loop at both termini without extending into neighbouring domains.

Another probable cause of the insolubility could be that the different domains interact with each other, lending stability to each other in an analogous way to the repeats within a repeating region,

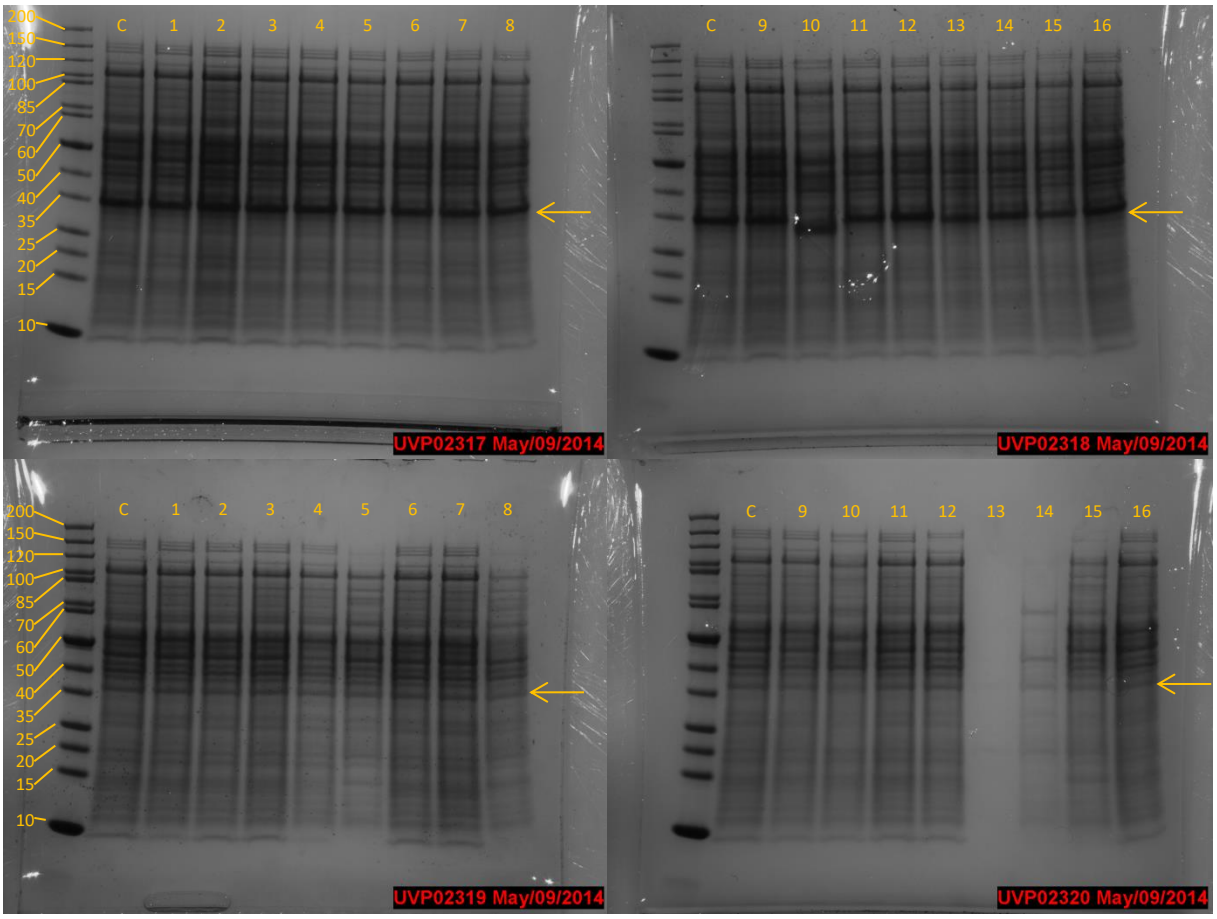


Figure 3-4 – Screening the effect of lysis additives on the solubility of my ANK construct. Visualised with white light after Coomassie blue staining. The top two gel photos show the total cell protein fractions and the bottom two show the total soluble protein fraction. The expected position is denoted by ←. Lanes C – Control; 1 – 0.75 M Trehalose; 2 – 1 M glycine betaine; 3 – 0.5 M mannitol; 4 – 0.1 M trisodium citrate; 5 – 1 M TMAO; 6 – 0.5 M L-proline; 7 – 1 M NDSB 195; 8 – 0.375 M L-arginine; 9 – 1 M xylitol; 10 – 0.01 M sodium selenite; 11 – 1 M NDSB 201; 12 – 0.1 M dipotassium hydrogen phosphate; 13 – 0.01 M copper (II) chloride; 14 – 0.5% (w/v) CTAB; 15 – 0.1% (v/v) SDS; 16 – (v/v) Triton X-100. The protein ladder was the same for all gels.

and separating them as I have done has led to them being insoluble. This could be solved by either co-expressing the constructs or by fusing my constructs back together by subcloning.

One final way to stabilise the insoluble constructs would be to express them in either a mammalian or insect system as these systems contain different protein chaperones which would be more similar to those found in human. The problem with this approach is that it is very difficult to scale up protein expression to the amounts obtainable in *E. coli*, and the sequence has already been codon-optimised for *E. coli* expression, likely leading to slower expression and lower overall yields in these systems. A possible solution to this would be to attempt expression in the *Pichia pastoris* yeast system, which can have comparable expression levels to recombinant expression in *E. coli*. However our research group does not have the facilities for culturing *P. pastoris* and the set up would have been unfeasible at this point.

3.5 Further Bioinformatics Studies of the Constructs

Before designing the extensions the protein sequences were analysed for their propensity to form repeat structure from the sequence alone and there are a number of online resources that are able to do this. Historically it has been difficult to predict repeating structure, due to the amount of variation that can exist around the consensus sequence, especially the detection of the terminal repeats which tend to be significantly different in composition but not tertiary structure.

A number of online repeat prediction programmes were first tested with proteins of known repeating structure. These test proteins were: β-catenin (UniProt ID P35222, PDB ID 2Z6H (Xing *et al.*, 2008)), an ARM repeat protein; gankyrin (UniProt ID O75832, PDB ID 1UOH (Krzywda *et al.*, 2004)), an ANK repeat protein; internalin B (UniProt ID P25147, PDB ID 2Y5Q (Ebbes *et al.*, 2011)), a

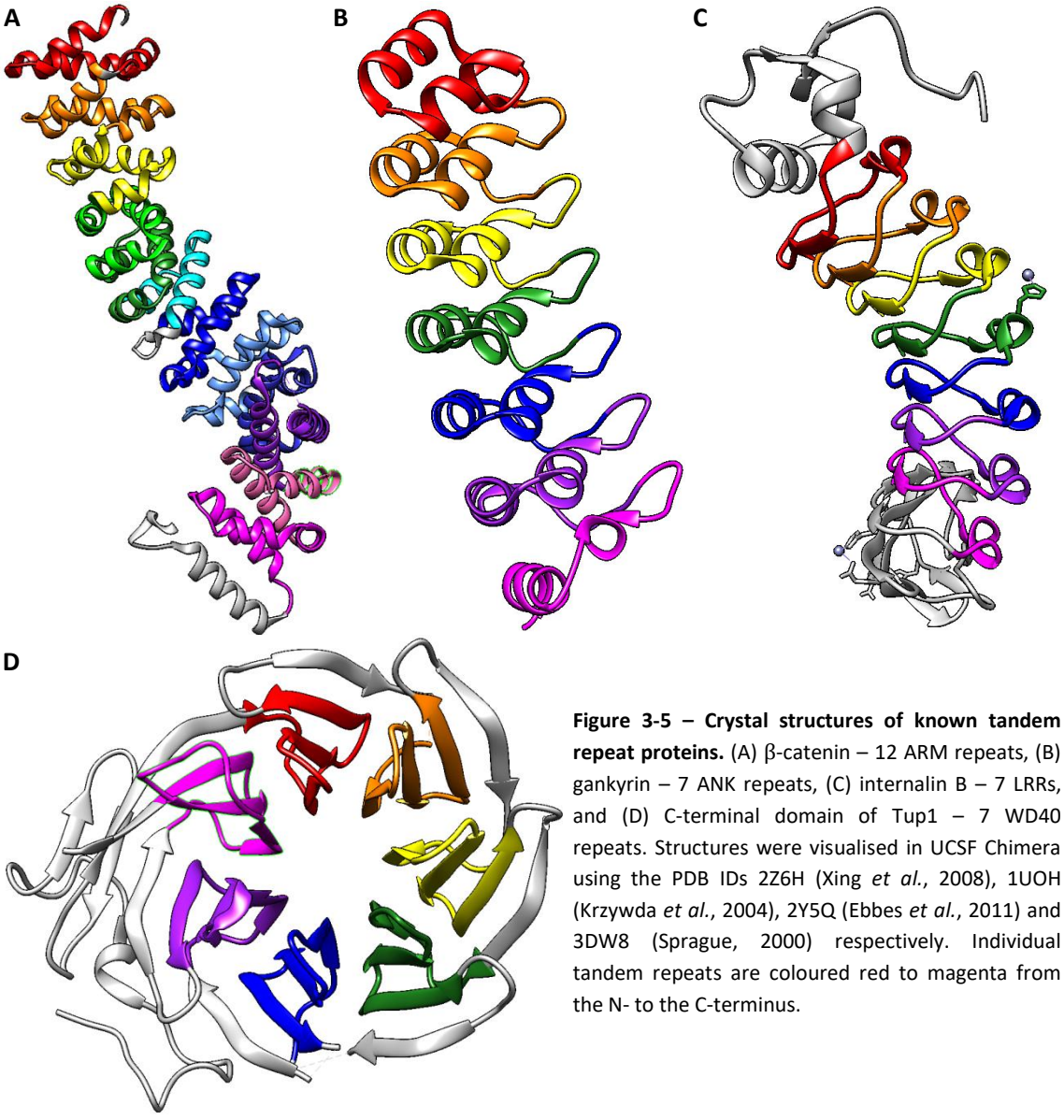


Figure 3-5 – Crystal structures of known tandem repeat proteins. (A) β-catenin – 12 ARM repeats, (B) gankyrin – 7 ANK repeats, (C) internalin B – 7 LRRs, and (D) C-terminal domain of Tup1 – 7 WD40 repeats. Structures were visualised in UCSF Chimera using the PDB IDs 2Z6H (Xing *et al.*, 2008), 1UOH (Krzywda *et al.*, 2004), 2Y5Q (Ebbes *et al.*, 2011) and 3DW8 (Sprague, 2000) respectively. Individual tandem repeats are coloured red to magenta from the N- to the C-terminus.

LRR repeat protein; and the C-terminal domain of Tup1 (UniProt ID P16649, PDB ID 1ERJ (Sprague, 2000)), a WD40 repeat protein (see Figure 3-5).

Three programmes were found to be fairly accurate: RADAR, which uses the RADAR algorithm (Heger and Holm, 2000; Goujon *et al.*, 2010); Motif Scan (Pagni *et al.*, 2007), which matches motifs against the 'motif' databases, Prosite (Sigrist *et al.*, 2013), Pfam (Finn *et al.*, 2014) and HAMAP (Pedruzzi *et al.*, 2013); and InterProScan, which uses multiples algorithms including BLAST and HMMER to match protein sequence to data from multiple public databases (see paper for complete list (Jones *et al.*, 2014)).

For these four test proteins, RADAR was able to identify all ARM and LRR repeats including the terminal ones, all ANK repeats except the terminal ones, but none of the WD40 repeats. It also shifted the frame of the repeat motif in each case. Motif Scan was able to find all ARM repeats, all ANK and LRR repeats except for the terminal repeats and some WD40 but tended to underestimate the actual length of the repeat motifs. InterPro identified all ARM and WD40 repeats, all but one of the LRR repeats and all but two of the ANK repeats. However it was not especially accurate as to the location or length of the repeats but made up for the lack of accuracy by identifying "repeat-like domains" extending beyond the actual repeats.

When my ARM construct was analysed, RADAR identified ten repeats in sets of four, three and three, compared to the thirteen predicted by Mills *et al.*, which were not aligned as predicted by Mills *et al.*. InterPro was unable to identify any repeats but did identify an area of ARM-like fold that encompassed the predicted repeat region. Motif Scan, on the other hand, only identified a single potential ARM repeat with low certainty (see Figure 3-6a).

Analysis of the predicted ANK domain was less conclusive. RADAR identified no repeating structure, Motif Scan identified a single potential ANK repeat with low certainty, and InterPro identified only the N-terminal half as being an ANK-like fold domain (see Figure 3-6b).

When the predicted LRR domain was analysed, RADAR identified twelve of the fourteen predicted which were aligned as predicted by Mills *et al.*, although again the frame was shifted. Motif Scan identified only six repeats but with high confidence and InterPro predicted eleven repeats, with the three spaces between the predicted repeats, being the size of repeats (see Figure 3-6c).

Analysis of the WD40 construct was also not particularly reassuring. RADAR identified some repeating structure but not anything resembling WD40 repeats and Motif Scan identified a single potential WD40 repeat with low certainty. However InterPro did identify that there was a WD40-like fold domain, encompassing most of the predicted repeat domain (see Figure 3-6d).

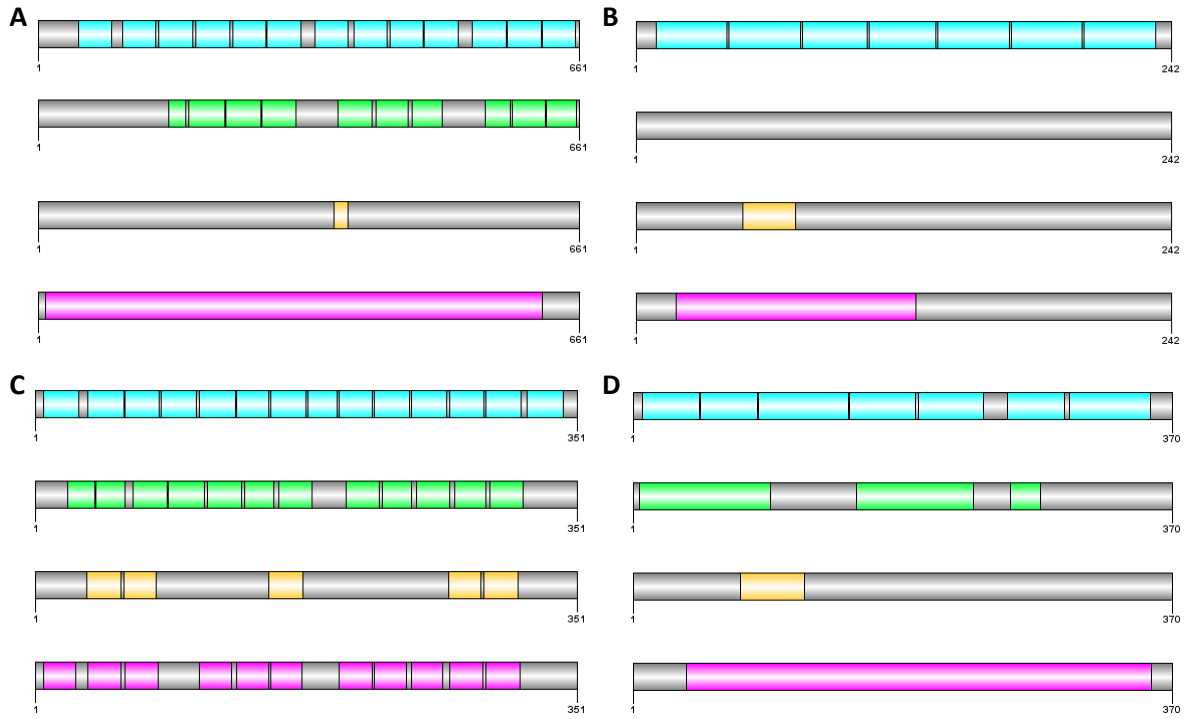


Figure 3-6 – Predicted repeat locations in the LRRK2 constructs. (A), (B), (C) and (D) are the ARM, ANK, LRR and WD40 constructs respectively. The positions of the repeats, as predicted by Mills *et al.* are denoted in blue. Predictions by RADAR, Motif Scan and InterProScan are marked in green, yellow and purple respectively.

Taken together these predictions suggest there is likely to be some inaccuracy in the predicted positions of the individual repeats and the repeat domains by Mills *et al.*, which led me to believe that extending the constructs may improve their solubility. Although tandem repeating structure has historically been difficult to predict, consensus sequences for these four motifs have been known for a while (Andrade *et al.*, 2001; Li and Roberts, 2001; Mosavi, Minor and Peng, 2002; Bella *et al.*, 2008). The three prediction tools I have used all try to predict structural elements from proteins sequence whereas the predictions of Mills *et al.* come from homology modelling. Both techniques have different strengths and pitfalls and so corroboration between these two predictive methods gave me some confidence in the existence of the proposed repeats for the ARM and LRR constructs, less so for the ANK and WD40 constructs.

3.6 Designing the Construct Extensions

The ARM construct already extends all the way to the N-terminus of LRRK2 so it could only be extended from the C-terminus. However I then encountered the problem that the loop between ARM and ANK is quite short. In my original construct design, I had chosen not to have any sequence overlap between the four constructs in order to study the domains in isolation. However in order to extend the ARM construct in the hope of increasing expression level and solubility I had to overlap with the ANK construct. When I designed the original constructs, a hydrophobicity plot of each construct and the connecting loops was produced in order to help determine the optimum construct

length. During this process it was noticed that the N-terminal region of the ANK construct is particularly hydrophilic and so would help aid in solubilising the ARM construct.

This overlap of the ARM and ANK construct, although not what I had originally intended, did allow me to plan ahead to future experiments. Silent mutation analysis (Palmer, 2014) was performed on the region where the two constructs would overlap and a potential unique BbsI restriction site was found. This restriction site would allow me to make a contiguous ARM-ANK domain fusion protein at a future point. The C-terminal ARM construct extension (see Appendix B, Section B.1.1) was designed to include this BbsI site as well as HindIII sites restriction sites at the 5' and 3' ends with which to subclone the extension. The stop codon at the end of the original ARM construct would need to be removed by RTH PCR afterwards, and the BbsI restriction site would need to be mutated into the ANK construct by SDM.

Extending the ANK N-terminally did not seem to be a good idea as the ARM-ANK loops is short and the aforementioned hydrophobicity plots had shown the C-terminal region of the ARM construct is not particularly hydrophilic. The ANK-LRR loop, however, is long and quite hydrophilic and so a C-terminal extension was designed for the ANK construct. It was not practical to design this extension to overlap directly with the LRR construct but as I intended to extend the LRR construct N-terminally, the C-terminal ANK extension was designed to overlap with the N-terminal LRR extension. Silent mutation analysis of the ANK construct revealed a potential AfeI restriction site in C-terminal region that would need to be introduced by SDM, so AfeI-HindIII was chosen for subcloning. Similar analysis of the ANK-LRR loop revealed a potential BbvCI restriction site which was chosen to facilitate crosslinking the ANK and LRR constructs (see Appendix B, Section B.1.3).

The LRR construct was extended N-terminally rather than C-terminally to avoid extending into the COR domain and because the ANK-LRR loop was more hydrophilic and would allow the ANK and LRR domains to be fused. A potential SpeI restriction site was found in the N-terminal region of the LRR construct and so BamHI-SpeI was chosen for subcloning, and BbvCI was already chosen for crosslinking (see Appendix B, Section B.1.4).

3.7 Subcloning the Extensions

The extension for the ARM construct was ordered as a double-stranded DNA string from Life Technologies' GeneArt. The C-terminal ARM extension, ARM construct in both pRSET and pRSET-GST, was digested with HindIII, then digested ARM construct in it vectors was then treated with Antarctic phosphatase before being combined with the digested gene string and ligated as described in Section 2.2.2. The ligated plasmids were then amplified as described in Section 2.2.5 then sent for

sequencing (Source Bioscience). The results from sequencing revealed that all of plasmids were re-ligated, original ARM construct.

This process was repeated twice more with no positive results, at which point it was realised that the strings as designed did not include excess bases on either side of the cut sites, and HindIII is a restriction enzyme that requires a minimum of three excess bases for effective cleavage (NEB, no date). So primers had to be designed and ordered to extend the extension to the point where they could be digested with HindIII (see Appendix B, Section B.1.2). Fortunately I had not yet ordered the extensions for the ANK and LRR constructs and so their design could be rectified before I made the same mistake.

The C-terminal ARM extension was extend using PCR, as described in Section 2.2.2, digested with HindIII, and ligated into the ARM constructs in each vector, as above. Again sequencing revealed no successful subcloning had occurred.

3.8 Co-expression of the ARM, ANK and LRR Constructs

In order to facilitate co-expression of my repeat domain constructs, an attempt was made to clone the ARM, ANK and LRR into the pET-41a(+) and pACYCDuet-1 plasmid vectors, which contain a kanamycin and a chloramphenicol resistance gene respectively. Each of the three constructs, along with the pET-41(+) and pACYCDuet-1 vectors, were digested with BamHI and HindIII then gel extracted, as described in Sections 2.2.2. The two vectors were both treated with Antarctic phosphatase, and each of three constructs were ligated into each vector, amplified by miniprep and sequenced. In this first attempt the ANK and LRR constructs were successfully cloned into the pET-41a(+) vector and the LRR construct was successfully cloned into the pACYCDuet-1 vector's first MCS.

However it was at this point that it was realised that in the pACYCDuet-1 vector's first MCS, the BamHI and HindIII restriction sites resulted in the LRR construct being in a different reading frame, as would also have been the case for my other constructs. This meant that I was unable to attempt the triple co-expression straight away but could later be rectified by RTH SDM to remove sufficient bases to bring the construct back into frame.

Trial co-expressions were attempted by transforming separate aliquots of C41 cells simultaneously with equal molar amounts of ARM in pGST and ANK in pET-41a(+), ARM in pGST and LRR in pET-41a(+), and ANK in pGST and LRR in pET-41a(+), and grown overnight on agar plates containing both ampicillin and kanamycin. Each of these four plasmids were also individually transformed into separate aliquots of C41 cells alongside these co-expressions. The single transformants were each

grown on agar containing ampicillin only, kanamycin only or both antibiotics under the same conditions to act as controls demonstrating the requirement of two plasmids for survival.

The only successful co-transformation from this test was ARM and LRR co-expression, which is the pair I had predicted to be the least likely to successfully co-express, as the two domains do not neighbour each other in the sequence and the ANK domain between them would likely physically separate the two in the full-length protein. There are two plausible explanations for this low success rate. Either during the resting period (1 hr at 37 °C) one of the two plasmids was ejected by the bacterium to reduce energy expenditure or the double transformation had not succeeded in the first place, resulting in 0 colonies formed.

Colonies from the ARM + LRR co-transformation were picked and inoculated into 5 ml of fresh 2xYT media containing both kanamycin and ampicillin for a protein expression test as described in Section 2.3. The cells were harvested and lysed with BugBuster, as described in Section 2.4, then the results were analysed by SDS-PAGE (see Figure 3-7).

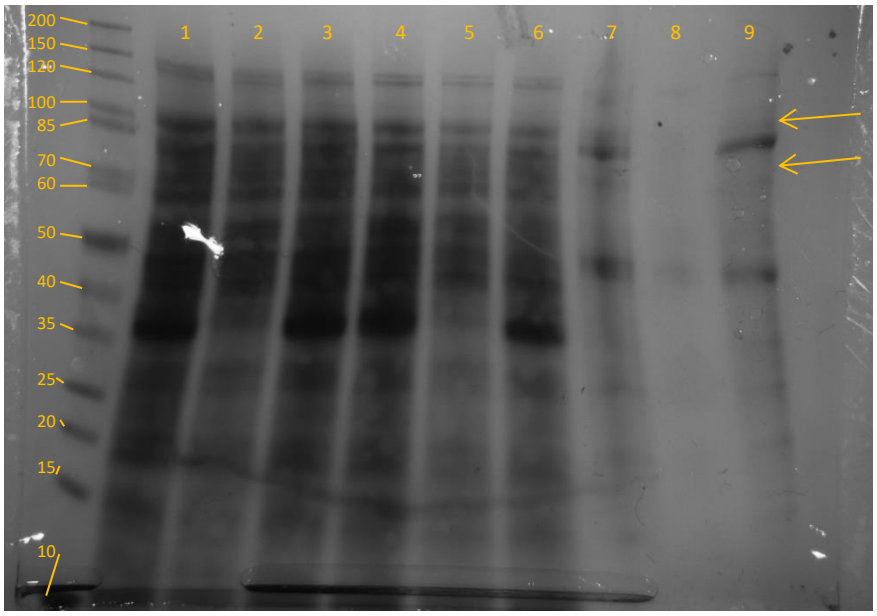


Figure 3-7 – SDS-PAGE analysis of the co-expression of ARM and LRR. Lanes 1-3 – Total cell protein for ARM only, LRR only and ARM +LRR; 4-6 – Total soluble protein for ARM only, LRR only and ARM +LRR; 7-9 – Total insoluble protein for ARM only, LRR only and ARM +LRR. The top ← indicates the expected position of the GST-ARM fusion protein (M_w ≈ 100.5 kDa) and the bottom indicates that of the GST-His-S-tag-LRR fusion protein (M_w ≈ 72.5 kDa).

From the pattern of bands, the ARM and the LRR constructs did not appear to co-express, with the expression pattern resembling that of the ARM only expression (see Figures 3-1 (F) and 3-8). I could be sure that both plasmids must still be present within the cells as they would have been unable to grow in the presence of both antibiotics otherwise. The likely explanation for this is a difference in plasmid copy numbers, pET-41a(+) is a low copy number plasmid (≈40) whereas pRSET derivatives are high copy number (>200). This would have resulted in different relative expression levels or even

preferential expression of one over the other. Ideally what I should have done is used a single plasmid with more than one MCS, such as the pACYCDuet-1 vector, as this would have resolved the issues I had with both the double transformation and the relative expression levels.

3.9 Mammalian expression system

The expression of each of my LRRK2 constructs in a mammalian cell system was also tested using a TNT® Quick Coupled Transcription/Translation kit (Promega). This kit uses rabbit reticulocyte lysate combined with additional recombinant proteins which allow *in vitro* translation of proteins under the control of either the T7 or SP6 promotors in a eukaryotic cell system. This kit was combined with FluoroTect™ $Green_{Lys}$ (Promega), an in vitro translation labelling system which uses lysine-charged tRNA that was already labelled with BODIPY-FL fluorescent dye at the ε amine group. Proteins expressed will incorporate fluorescently labelled lysines during in vitro translation. The TNT® T7 Quick Master Mix also contains a plasmid encoding for luciferase under the T7 promoter which acts as a positive control for transcription/translation. In this system, the luciferase will also be fluorescently labelled by the FluoroTect™ $Green_{Lys}$. These two kits were used according to their manufacturer's instructions and the reaction mixture was as described in Table 3-2.

Table 3-2: Reaction mixture for eukaryotic expression using TNT Quick Coupled Transcription/Translation kit

TnT Mammalian Transcription/Translation	
8 µl	TNT® T7 Quick Master Mix
0.2 µl	1 mM L-Methionine
1 µl	Plasmid DNA at ≈ 100 ng μl^{-1}*
0.4 µl	FluoroTect™ $Green_{Lys}$
0.4 µl	mQ water
10 µl	Total Volume

*Actual plasmid concentrations used were 105.4 ng μl^{-1}, 81.5 ng μl^{-1}, 114.3 ng μl^{-1}, and 97.5 ng μl^{-1} for the ARM, ANK, LRR and WD40 constructs respectively.

A reaction was set up for each of my four GST-tagged LRRK2 constructs as well as a plasmid-free control (1 µl of mQ water was used instead of plasmid), incubated for 90 min at 37°C and then analysed by SDS-PAGE. BenchMark™ Fluorescent Protein Standard (ThermoFisher Scientific) was used in place of the usual protein ladder and the gel was fixed using 25% (v/v) isopropanol, 10% (v/v) acetic acid in mQ water prior to being visualised under UV-light (see Figure 3-8). From the SDS-PAGE analysis it appeared that none of the constructs expressed in the mammalian system. This result is likely due to the plasmid DNA sequence being codon optimised for recombinant expression in *E. coli*.

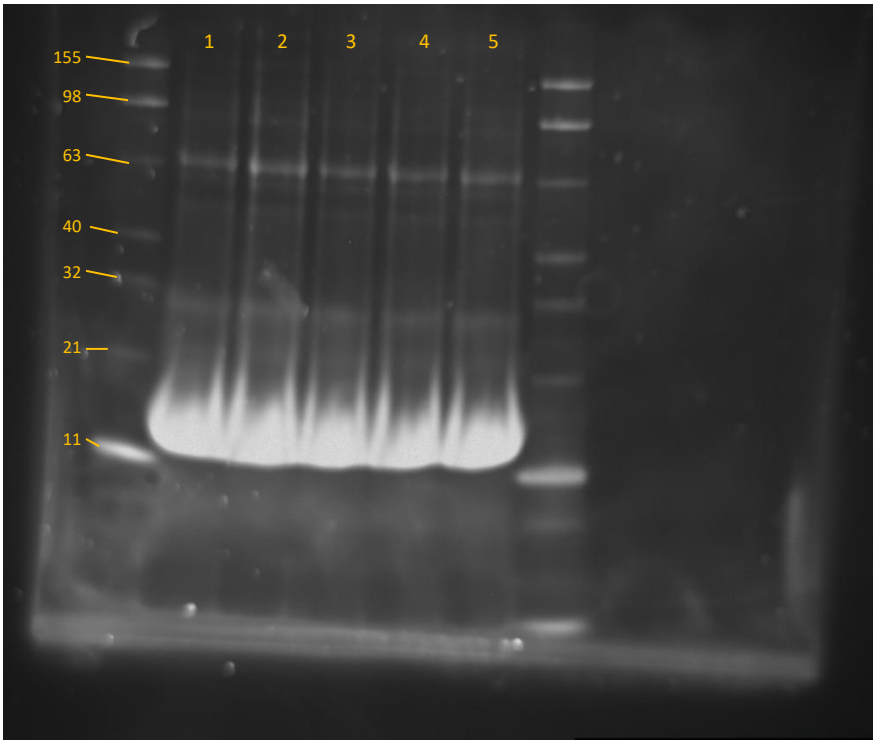


Figure 3-8 – SDS-PAGE analysis of the mammalian expression of my LRRK2 constructs. Visualised with UV light. Lane 1 – No plasmid control; 2 – ARM; 3 – ANK; 4 – LRR; 5 – WD40. A ubiquitous band can be seen at MW ≈ 63 kDa around the expected MW for luciferase (60.5 kDa), which was taken as evidence that the positive control was successful.

3.10 Refolding from Inclusion Bodies

Lastly, an attempt was made to produce soluble protein by refolding insoluble protein from inclusion bodies (Singh and Panda, 2005). This was a far from ideal solution and very much a last ditch attempt to express my LRRK2 constructs solubly. Partially folded and aggregated proteins sequestered in inclusion bodies are denatured and solubilised using chaotropic agents, e.g. urea or GnHCl, then the chaotropic agent is removed to allow the re-solubilised proteins to refold. Refolding from inclusion bodies was avoided previously, as it is necessary to demonstrate that protein purified by this method refolds back into the native structure. However, at the time these experiments were performed, the crystal structures for this region of LRRK2 was (and at the time of writing still is) unsolved and therefore demonstrating that the protein has refolded back to the native state would not be possible. Nevertheless I did attempt to produce soluble protein by inclusion body refolding, as described below.

A 6 l prep of GST-ANK was made as described in Section 2.3 and lysed as described in Section 2.5.1. After removing the supernatant containing the soluble proteins, the cell pellet was resuspended in 100 ml of a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM DTT and 8 M urea, vortexed until homogenous and incubated at 4°C on a nutator for 1 hr. The cell debris was then re-pelleted by centrifuging at 35000 g at 4°C for 35 min, and the supernatant containing the solubilised inclusion body proteins was decanted.

In order to refold the extracted inclusion body proteins, the urea needs to be removed or diluted to a concentration where the folded state is favoured, usually below 50 mM urea (160-fold dilution). There are three main methods to do this, the first two of which involve diluting out the urea, either by dialysis, or by rapid dilution, each of which has problems associated with them. With rapid dilution the danger is that there is a kinetic trap along the folding pathway causing the protein to misfold upon renaturing. For dialysis kinetic traps are avoided although other time-dependent mechanisms of aggregation are more like to occur. The third method is by affinity purification, the target protein is extracted from solution via its existing purification tag. However this is not possible in this case as the GST-fusion purification tag is also denatured in the presence of 8 M urea and unable to bind to glutathione resin. Had I chosen a His-tagged construct, then nickel-affinity purification would have been possible.

18 ml of decanted supernatant and sealed in 10 kDa MWCO dialysis tubing and dialysed into 3 l of pre-cooled 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM DTT at 4°C for 16 hrs (>165-fold dilution). Meanwhile another 12 ml of decanted supernatant was mixed rapidly into 2 l of the same buffer, also precooled and at 4°C. Upon doing so a white precipitate of protein immediately formed. Similarly upon returning to the dialysis it was found that the dialysis tubing was filled with a white precipitant. I took this as evidence that the GST-ANK construct was simply not soluble.

3.11 Concluding the LRRK2 Project

In this chapter I attempted to study the four tandem repeat domains of LRRK2, as previously proposed by Mills *et al*. (2012). However despite the application of many common biochemical methods to increase protein solubility I was unable to obtain soluble protein constructs comprising the proposed domains.

Although it must be admitted that not every method was attempted with each construct, and the techniques were not used in combination with each other, the fact that no single technique resulted in a significant increase in soluble expression of any domain, strongly suggests that there are other factors significantly affecting the protein stability. It is likely that the neighbouring domains provide extensive stabilising interactions and it is their absence that resulted in this significant loss of stability. Although these experiments are not sufficient to either prove or disprove the predictions of Mills *et al.*, I believe they are sufficient to demonstrate that recombinant expression of the isolated domains in *E. coli* is not the optimal way to study this protein.

It may still be possible to study these domains in isolation if their codons were optimised for expression in other systems e.g. baculovirus/insect cells (Kost, Condreay and Jarvis, 2005), which

utilise different chaperone proteins to aid soluble expression. However, I expect that others may find more success in studying these proposed domains using larger constructs comprising multiple domains.

Although crystallisation of multi-domain proteins with intrinsically disordered linkers for x-ray diffraction is often difficult to optimise if not sometimes impossible, there are now many other techniques for determining protein structure such as cryo-electron microscopy (cryoEM) and atomic force microscopy (AFM) (lower resolution) (Ozkan *et al.*, 2016; Cuniasse *et al.*, 2017; Ott *et al.*, 2017; Perilla *et al.*, 2017).

Regarding AFM, our collaborator Dr. Pia Jeggle (Edwardson group, Department of Pharmacology, University of Cambridge) expressed full-length, myc-tagged, human LRRK2 in a human cell line and was able to obtain AFM images of the protein. However this is not an area of research that the Itzhaki group has much expertise in, and so I did not pursue this avenue of investigation any further and instead refocused my studies on a new topic: β-catenin and TCF7L2.

Chapter 4 – β-Catenin and TCF7L2: Method Development

As detailed in Chapter 1, β-catenin is a cancer-associated ARM repeat protein of interest to the Itzhaki group. It interacts with a large number of partner proteins many of which are intrinsically disordered or have intrinsically disordered regions which form the binding interaction. One of these partners, which will be the subject of the remainder of this thesis, is the transcription factor, Transcription Factor 7-Like 2 (TCF7L2). TCF7L2 is often referred to as Tcf4, being the fourth identified member of the Tcf/LEF family of transcription factors, but I will be using TCF7L2 so as to avoid confusion with TCF4 (Transcription Factor 4).

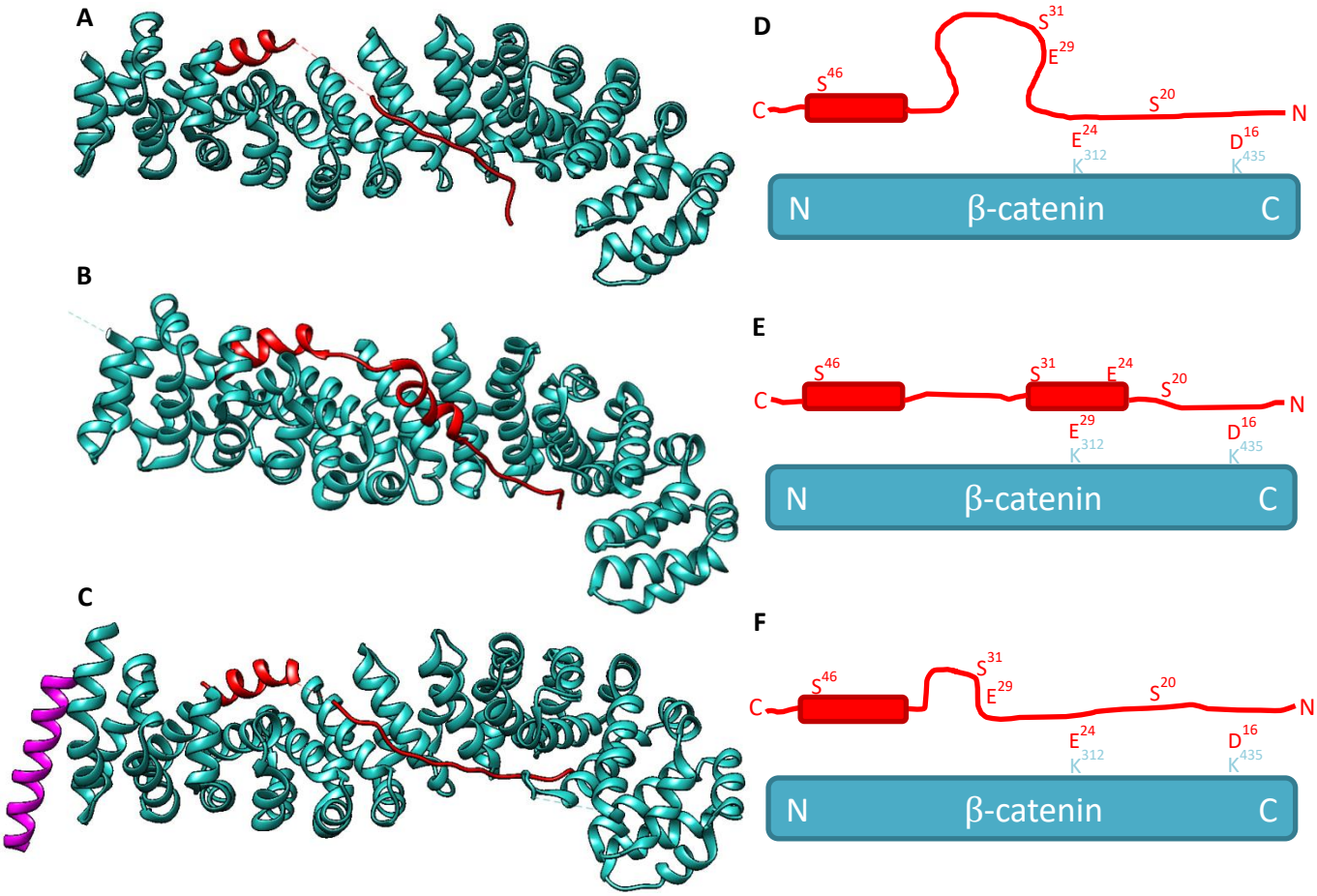


Figure 4-1 – The crystal structures of the β-catenin/TCF7L2 complex. (A), (B) and (C) are the crystal structures of TCF7L2 (red) in complex with β-catenin (cyan) obtained by Poy *et al.*, Graham *et al.* and Sampietro *et al.* respectively. In structure (C), BCL9 (purple) is also shown in complex with β-catenin. (D), (E) and (F) are cartoon diagrams of (A), (B) and (C) respectively, demonstrating the differences between the three structures and highlighting the relative positions of significant residues. Structure images (A) (B) and (C) were created using UCSF Chimera (Pettersen *et al.*, 2004) from PDB IDs 1JPW (Poy *et al.*, 2001), 1JDH (Graham *et al.*, 2001) and 2GL7 (Sampietro *et al.*, 2006) respectively.

The origin of this project is a curiosity in the crystal structures of human TCF7L2 binding to β-catenin. There exist three unique crystal structures of TCF7L2 bound to β-catenin in the PDB and they show significant differences between them (see Figure 4-1). I noted in Section 1.2.1, that the IDR of the tumour suppressor protein p53 is capable of adopting different structural conformations when binding to different partner, allowing a diversity of function. Similarly I discussed in Sections 1.5.2 and 1.6, the TCF7L2-β-catenin complex recruits a number other proteins in order to perform its functions. This then raises the question: are the alternative conformations seen in the crystal structure biologically relevant?

For the remainder of this thesis I will be investigating the biophysical origins of this difference and the potential consequences of forming different structures in complex including the mechanisms of binding and unbinding. In this chapter I will discuss the development of a protocol for the expression and purification of the N-terminal β-catenin-binding domain of wild-type TCF7L2. Additionally I describe the development of a fluorescence reporter system for the TCF7L2-β-catenin interaction which was required for the work described in later chapters.

4.1 β-Catenin and TCF7L2 Purification

The β-catenin construct used in this thesis was provided by Professor W. Weis, hereafter referred to as Weis β-catenin. It comprises residues 134-671, which corresponds to the ARM repeat domain as described by Choi, Huber and Weis (2006) (see Appendix A, Section A.2). The majority of the methods and techniques for the purification, storage and handling of Weis β-catenin, as described in Chapter 2, were devised and optimised by Dr. S.D. Dunbar and Dr. P.J.E. Rowling, who were working with Weis β-catenin before I began this project and for which I am very grateful.

In the three different crystal structures of TCF7L2 bound to β-catenin, each paper uses a slightly different TCF7L2 construct: Poy *et al.* uses residues 8-54, Graham *et al.* uses 1-54 and Sampietro *et al.* uses 1-53. I opted to use the 1-54 construct as it included all residues used in these structures (see Appendix A, Section A.3.1). This sequence for this was taken from UniProt's Protein Knowledgebase, ID Q9NQB0, codon optimised for expression in *E. coli*, ordered as a gene string from ThermoFisher Scientific and cloned into the expression plasmid vectors: pRSET and pGST (see Section 2.2.2).

Initial test expressions of TCF7L2 showed that the majority of TCF7L2 expressed in the soluble fraction when expression was induced with 500 µM IPTG and incubated at 25°C overnight. TCF7L2 was then grown and expressed on a large scale, as described in Section 2.3, and purified using a buffer system based on 50 mM potassium phosphate at pH 6.0, 150 mM NaCl and 1 mM DTT. This

buffer was chosen as the TCF7L2 had a pI of 3.88 as determined by ProtParam (Gasteiger *et al.*, 2005). This was later replaced by buffers based on 1x PBS pH 7.4 and 1 mM DTT when it was decided that experiments were to be done in that buffer, thereby avoiding extra buffer exchange steps.

In my first attempt to purify TCF7L2, I used a method that Dr. P.J.E. Rowling had developed to purify other small intrinsically disordered proteins (IDPs) which she was working on. After the batch affinity purification step (Section 2.5.2), the few remaining protein contaminants were found to be sufficiently different in molecular weight from that of my TCF7L2 construct that they would theoretically be separable using a Vivaspin centrifugal concentrator (GE Healthcare) of a suitable molecular weight cut off (MWCO). For my wild-type TCF7L2 (1-54) construct, which has a molecular weight of 5.9 kDa after thrombin cleavage, a 10 kDa MWCO Vivaspin followed by a 3 kDa MWCO Vivaspin should remove all protein contaminants, and have the added benefit of concentrating the protein as required.

However upon testing this hypothesis, it was found that this construct not only did not pass through a 10 kDa MWCO filter it also would not pass through a 30 kDa MWCO filter, and therefore not separate the TCF7L2 from other larger contaminants. It is common for IDPs to behave as though they had a higher molecular weight (Gontero and Maberly, 2012). This is due IDPs having a higher hydrodynamic radius for their mass than globular proteins as a result of their lack of tertiary structure (English *et al.*, 2016; Tomasso *et al.*, 2016). Many common lab techniques such as SDS-PAGE actually separate proteins by hydrodynamic radii, which for globular proteins are proportional to molecular mass, so it is likely that the MWCOs stated by the manufacturer are calibrated for globular proteins.

In further trial purifications it was found that the wild-type TCF7L2 (1-54) construct does not bind to a MonoQ column in either 50 mM KPhos pH 6.0/150 mM NaCl/1 mM DTT or 1x PBS pH 7.4/1 mM DTT. However it was possible to isolate TCF7L2 using a G75 SEC column and so this was used in all future preparations of TCF7L2 and mutation variants. TCF7L2 was obtained in sufficient quantity that I felt that no further optimisation was required.

4.2 Labelling with FITC

The wild-type TCF7L2 construct contains no tryptophan or tyrosine residues, and therefore shows no intrinsic fluorescence. This presents a problem, as the majority of biophysical techniques employed in this Thesis use fluorescence as a way of monitoring the progress of a reaction, and ideally I would have liked to have used the intrinsic fluorescence of my proteins. The Weis β-catenin does contain three tryptophan and ten tyrosine residues; however Dr. P.J.E. Rowling and Dr. S.D. Dunbar had

found previously that there is not a sufficiently large change in fluorescence on β-catenin binding to other protein partners to be of use, likely due to their buried nature.

The main technique available for monitoring the binding equilibrium of proteins, without using fluorescence, is isothermal titration calorimetry (ITC), which monitors the changes in heat, either released or absorbed in a reaction. However the VP-ITC requires large quantities of both binding partners and, despite a significant amount of work by Drs. Rowling and Dunbar, β-catenin expression was not high enough for the technique to be used throughout this project. Consequently I opted to label my TCF7L2 (1-54) constructs with commercial fluorescent dyes, although some ITC was done to ensure the labelling process did not significantly affect binding.

Initially I chose to label the wild-type TCF7L2 (1-54) with fluorescein isothiocyanate (FITC) as this construct contains no cysteine residues but it does contain three lysine residues: K^{22}, K^{30} and K^{45}. However after several labelling reactions it was found that labelling wild-type TCF7L2 (1-54) with FITC was not consistent, achieving between 1.8 and 2.5 fluorescein moieties per TCF7L2 on average. Also on further inspection of the crystal structures it was noted that two of the three lysine residues, K^{22} and K^{45}, are located at the binding interface between β-catenin and TCF7L2, and was therefore likely to affect binding.

This lack of consistency was undesirable so I tried an alternative labelling method. The wild-type construct contains no natural cysteine residues so I decided to engineer one in by SDM. Doing so would mean that the construct would be less native-like and my results would be less representative of the β-catenin-TCF7L2 *in vivo* interaction, however it would give me more precise control over the labelling reaction.

4.3 Labelling with Fluorescein Maleimide and the Choice of Labelling Position

To minimise the effect of the SDM on TCF7L2, I chose to make a conservative serine to cysteine mutation in the wild-type TCF7L2 (1-54). I also considered introducing a tryptophan residue into the construct, and to use its intrinsic fluorescence to monitor the binding reaction. However one tryptophan residue would likely not produce a strong signal change to be observed over the background signal from β-catenin and was likely to have a significant impact on the expression and solubility of the construct.

The wild-type TCF7L2 (1-54) construct contains eight serine residues, excluding the one left behind as a result of removing the purification tag with thrombin. The crystal structures were inspected to

find a position where the addition of a bulky, hydrophobic dye molecule would not have a significant steric effect on binding but would be able to report on the different observed crystal structures. To aid in this a contact map analysis was performed on each of the crystal structures (Sobolev *et al.*, 2005).

A

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Residue	M	P	Q	L	N	G	G	G	G	D	D	L	G	A	N	D	E	L	I	S	F	K	D	E	G	E	Q	E	E	K	S	S	E	N	S	S	A	E	R	D	L	A	D	V	K	S	S	L	V	N	E	S	E	T
Poy *et al.*								UNSEEN					Extended																UNSEEN																								α-Helix					UNSEEN			
Graham *et al.*	UNSEEN												Extended								α-Helix 1														Extended								α-Helix 2										UNSEEN			
Sampietro *et al.*	UNSEEN												Extended																		UNSEEN																		α-Helix								UNSEEN			

B

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Residue	M	P	Q	L	N	G	G	G	G	D	D	L	G	A	N	D	E	L	I	S	F	K	D	E	G	E	Q	E	E	K	S	S	E	N	S	S	A	E	R	D	L	A	D	V	K	S	S	L	V	N	E	S	E	T
Poy *et al.*								0	0	0	0	0	1	2	4	7	6	5	7	1	8	2	4	6	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	7	0	2	5	6	1	1	7	4	0	0	0	0	0	
Graham *et al.*	0	0	0	0	0	0	0	0	0	0	0	0	4	5	6	5	7	6	4	6	1	9	0	0	2	3	1	0	2	5	0	0	1	2	4	4	3	4	4	1	3	7	0	2	5	6	1	1	7	4	0	0	0	0
Sampietro *et al.*	0	0	0	0	0	0	0	0	0	0	0	0	6	6	6	3	7	8	4	6	1	9	5	5	6	3	3	3	4	0	0	0	0	0	0	0	0	5	0	3	7	0	1	5	6	1	1	7	4	1	0	0	0	

Figure 4-2 – Structure maps of the three β-catenin/TCF7L2 crystal structures. (A) is a map of how major structural elements differ between the three crystal structures, by residue. (B) is a map of the number of contacts each residue makes with β-catenin as determined by contact map analysis (Sobolev *et al.*, 2005) with the default range and no minimum threshold for contact area.

The contact map analysis revealed that only one serine residue; S^{31}, formed no contacts to β-catenin in any of the crystal structures, indicating that it is in a sterically unhindered position. This position also has the additional benefit of being in different environments between the E^{24}-bound structures (Poy *et al.* and Sampietro *et al.*) and the E^{29}-bound structure (Graham *et al.*). In the E^{24}-bound structures, S^{31} is in a disordered unresolved loop, whereas, in the E^{29}-bound structure, it is in the first α-helix, pointing away from the surface of β-catenin. Consequently S^{31} was chosen for use as a controlled labelling position. The TCF7L2 (1-54) S31C construct is hereafter referred to as the "WT" S31C, as it is used as a pseudo wild-type with which all other TCF7L2 constructs will be compared (see Appendix A, Section A.3.2).

4.4 Testing effect of label using ITC

In order to determine whether the S31C mutation, and its subsequent labelling with fluorescein maleimide, had had a significant effect on TCF7L2 binding to β-catenin, equilibrium binding experiments were performed using ITC. These ITC experiments were performed as described in Section 2.7. Previous studies of TCF7L2-β-catenin binding have used ELISA (Omer *et al.*, 1999) and ITC (Sun and Weis, 2011) to determine the K_d of the interaction using similar constructs. The values they obtained were 15 ± 6 nM and 16 ± 3 nM for TCF7L2 (1-53) and TCF7L2 (1-57), respectively.

The results of these ITC experiments are shown in Figure 4-3 and are summarised in Table 4-1. As there was no significant difference between my wild-type TCF7L2 (1-54), "WT" S31C, fluorescein-labelled "WT" S31C ("WT" flS31C), and the literature values, I felt I could proceed in using the "WT" flS31C construct as a substitute for true wild-type, TCF7L2 (1-54). These experiments were only performed once as they were intended to be a "proof of concept" that the addition of a fluorescent

label at position 31 did not significantly affect binding. Going forward with the project I opted not to use ITC any further in this study, as it was too consuming of resources, i.e. β-catenin, which, as already mentioned, was proving difficult to optimise further for increased protein yield.

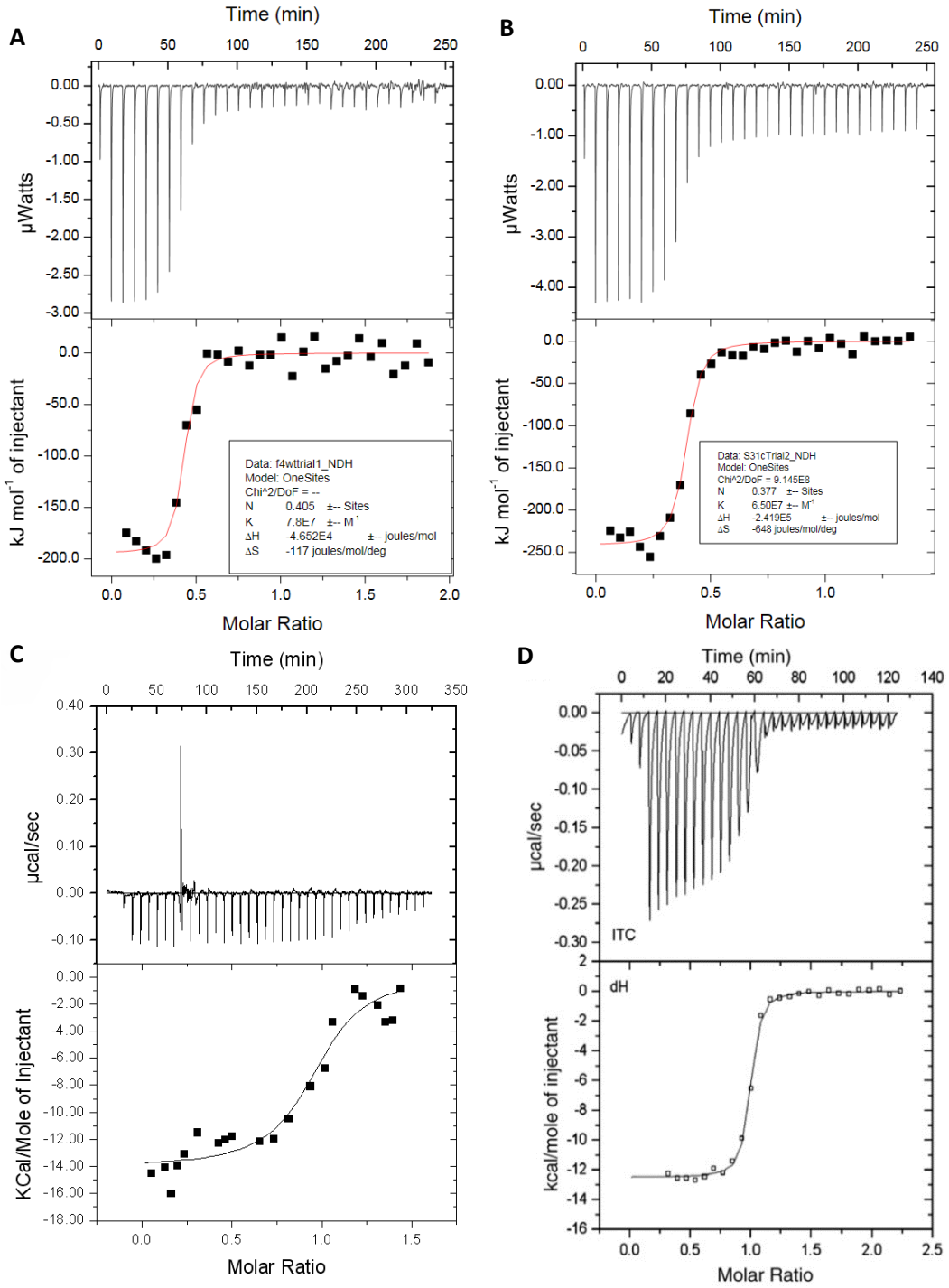


Figure 4-3 – ITC measurements for TCF7L2 constructs binding to β-catenin. The top panels show the heat signal obtained from a series of injections of different TCF7L2 into the ITC cell containing β-catenin, and the bottom panels show the binding curves calculated using the One-Site fitting model using the Origin software package. (A), (B) and (C) are the TCF7L2 (1-54), "WT" S31C and the fluorescein-labelling "WT" S31C construct binding to my Weis β-catenin construct respectively and (D) is data taken from (Sun and Weis, 2011) of TCF7L2 (1-57) binding to full length β-catenin. (A), (B) and (C) were performed in 25 mM Tris-HCl pH 8.9, 100 mM NaCl, 2 mM DTT at 30°C, whereas (D) was performed in 25 mM Tris pH 8.8, 100 mM NaCl, 2 mM DTT at 30°C.

Table 4-1: The K_d for different TCF7L2 constructs binding to β-catenin using ITC

Construct	Method	K_d (nM)	n[†]
TCF7L2 (1-53) (Omer *et. al.*)	ELISA	15 ± 6	3-20 not specified
TCF7L2 (1-57) (Sun & Weis)	ITC	16 ± 3	2
TCF7L2 (1-54)	ITC	13 ± 7	1
"WT" S31C	ITC	15 ± 4	1
"WT" flS31C	ITC	28 ± 12	1

[†] n is the number of repeats

4.5 Kinetic and Equilibrium Studies of "WT" flS31C

Having shown that the "WT" flS31C construct had suitably similar K_d to wild-type TCF7L2 (1-54) with ITC, I then wanted to switch to using other techniques to study the TCF7L2-β-catenin binding reaction. It was also at this time that I decided to switch to using the 1xPBS + 1 mM DTT buffer in order to be able to compare my results with those Drs. Rowling and Dunbar obtained for other β-catenin binding partners.

Fluorescence stopped-flow spectroscopy was used to determine the association rate constant of the "WT" flS31C construct in both the PBS buffer and the SunWeis Buffer (25 mM Tris pH 8.9, 100 mM NaCl, 2 mM DTT), as described in Section 2.8. For both buffers this produced a time-dependent decay in fluorescence intensity that was well described by the one-phase exponential function, equation [2-3]. In both cases, the two-phase exponential function, equation [2-3], sometimes seemed to produce a better fit, however GraphPad Prism 6 was not always able to fit two rate constants for a given concentration of β-catenin, and even when it could, there was a large error on the values obtained and there seemed to be no pattern to the second rate constant as the concentration of β-catenin was varied (see Figure 4-4).

Similarly, an attempt was made to determine the dissociation rate constant using stopped-flow spectroscopy, as described in Section 2.8. While a time-dependent increase in fluorescence intensity was observed the rate of change was too slow to be accurately measured, and on the longest practical recording time the trace never reached a plateau. I was able to determine the dissociation rate constant of "WT" flS31C in both PBS and SunWeis buffer, by using time-dependent fluorescence spectroscopy, as described in Section 2.9. Again the time-dependent increase in fluorescence intensity was observed but I was then able to record until a stable plateau had been reached. This time there were clearly two phases in the dissociation traces in both buffers, a major slow phase and a minor fast phase, as opposed to the single phase seen for association (see Figure 4-5).

The weighted averages of the association and dissociation rate constants for "WT" flS31C in both PBS and SunWeis buffer are summarised in Table 4-2. The errors given in the table are the weighted errors of the standard errors of the individual fits, and the true error is likely to be higher than those given. As all the kinetic rate constants for the two buffers are within error of each other I felt confident in continuing, using the PBS buffer. However it is worth noting the difference between the calculated kinetic K_d, $K_{d,kin}$, and the values found in the literature and from my own ITC experiments.

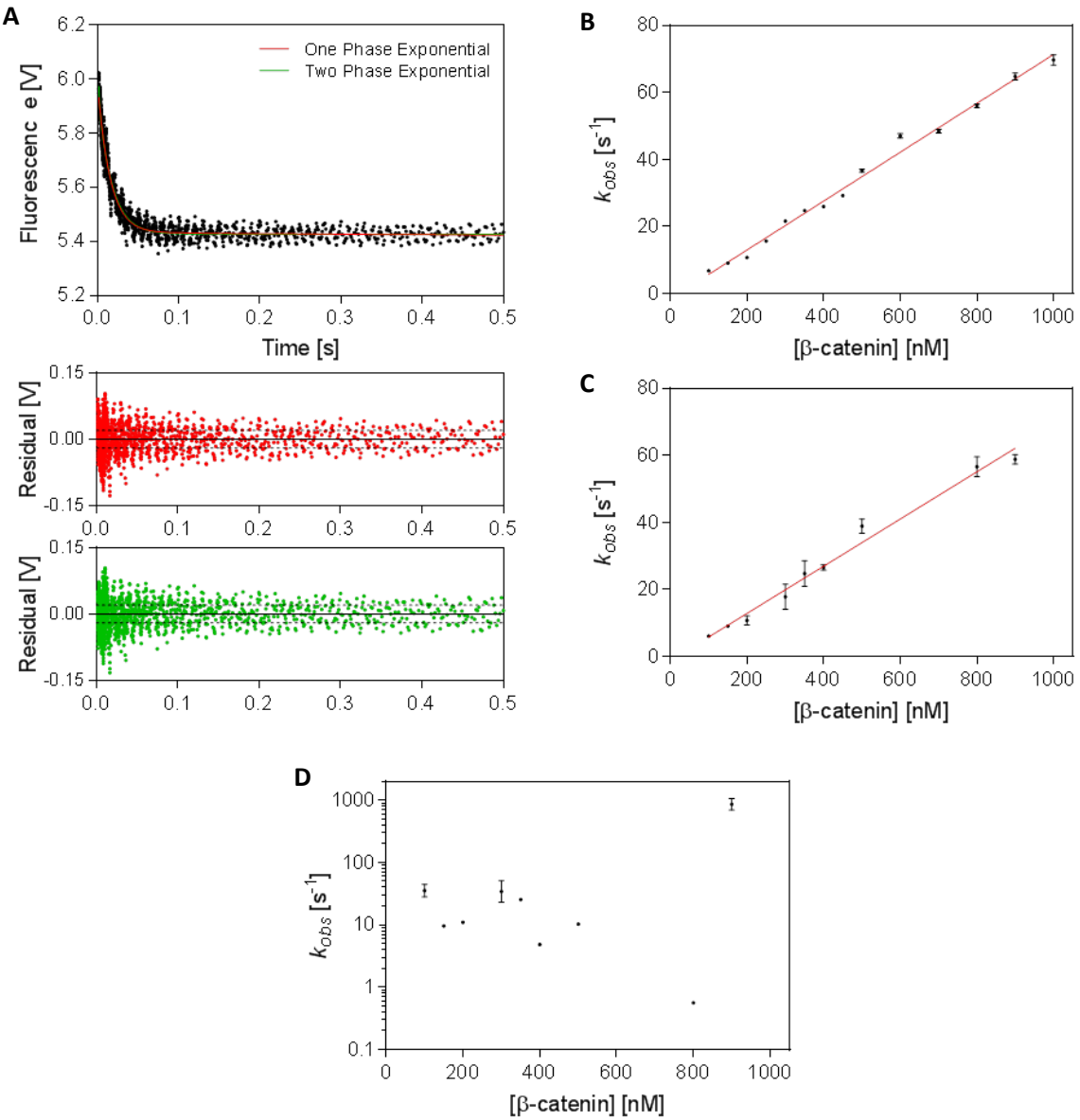


Figure 4-4 – Stopped-flow spectroscopy measurement of "WT" flS31C associating to β-catenin. (A) shows the time-dependent change in fluorescence when 10 nM "WT" flS31C and 900 nM β-catenin are mixed, taken from the same data set as (B), (C) and (D). These data were fitted with a one-phase (red) and a two-phase (green) exponential function and their associated residual traces are shown below. (B) is the plot of the observed one-phase rate constant against the concentration of β-catenin from which the association rate constant, k_{on}, is calculated. (C) and (D) are plots of the observed major and minor rate constants (respectively) against the concentration of β-catenin, when the same data used to plot (B) was instead fit with a two-phase exponential function. The errors in (B), (C) and (D) are the standard deviations of their respective fits. In the case of (D) a number of points had an error too large to be plot on the axes shown. In this case, the buffer was 1x PBS, 1 mM DTT at 15°C.

Table 4-2: Association and dissociation rate constants and $K_{d,kin}$ for "WT" fIS31C in different buffers.

Buffer	Association		Dissociation				$K_{d,kin}$ (pM)*
	k_{on} (x10^7 M^{-1} s^{-1})	n†	$k_{off,major}$ (x10^{-4} s^{-1})	$k_{off,minor}$ (x10^{-3} s^{-1})	% fast	n†	
SunWeis	6.11 ± 0.41	1	5.07 ± 0.12	1.19 ± 0.11	34 ± 5	4	8.3 ± 0.6
PBS	7.33 ± 0.14	16	5.73 ± 0.40	1.52 ± 0.28	23 ± 3	8	7.8 ± 0.6

* The kinetic K_d defined here as $K_{d,kin} = k_{off,major}/k_{on}$

† n is the number of repeats

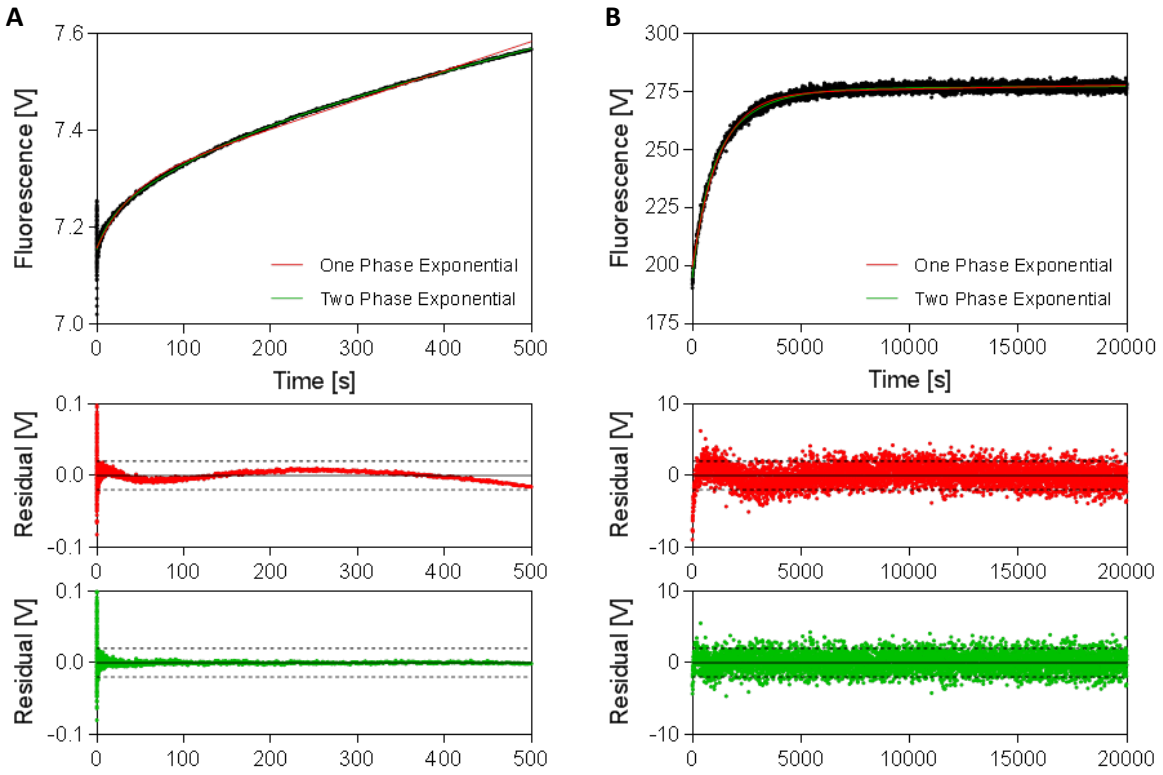


Figure 4-5 – Stopped-flow and time-dependent fluorescence spectroscopy measurements of the "WT" fIS31C-β-catenin complex dissociating. 200 nM "WT" fIS31C-β-catenin complex dissociating in the presence of 50 times molar excess of unlabelled WT TCF7L2 (1-54) as monitored by (A) stopped-flow spectroscopy and (B) time-dependent fluorescence spectroscopy. Both traces were fit with both a one-phase and a two-phase exponential function and the corresponding residuals are plotted below. Both experiments were performed in 1x PBS, 1 mM DTT at 15°C.

The K_d of an protein-protein or protein-ligand interaction can be calculated from its kinetic parameters using the simple relationship, $K_d = k_{off}/k_{on}$. In the case of association and/or dissociation being multi-phasic, it is conventional to use the major phase to calculate the kinetic K_d, $K_{d,kin}$, providing the major phase contributes significantly more to binding than any other phase. When this approach is applied to the kinetic data I obtained, the calculated $K_{d,kin}$s were three orders of magnitude stronger binding than obtained from ITC measurements, both my own and from the literature.

In order to investigate this difference, I used fluorescence anisotropy titration to study the system at equilibrium. Concentrated β-catenin was injected into a known concentration of "WT" fIS31C and

the change in fluorescence anisotropy measured, as described in Section 2.10. Fluorescence anisotropy is typically chosen over fluorescence intensity as it is concentration-independent. The fluorescence anisotropy was observed to increase with increasing β-catenin concentration until saturation of ligand. These traces were fit with three fairly common models of protein-ligand binding, to try to find the best approximate model: polarisation, accounting for ligand depletion (Equation [2-8]); the Hills equation (Equation [2-9]): and the MWC model (Equation [2-10]). However there was a general inconsistency in the traces, both in traces collected on the same day and between sets of traces collected on different days (see Figure 4-6). For these traces, the polarisation equation produced K_ds ranging from 3 pM to 700 pM, the Hills equation gave K_ds ranging from 2 nM to 7.5 nM, and the MWC gave K_ds ranging from 5 pM to 1000 pM.

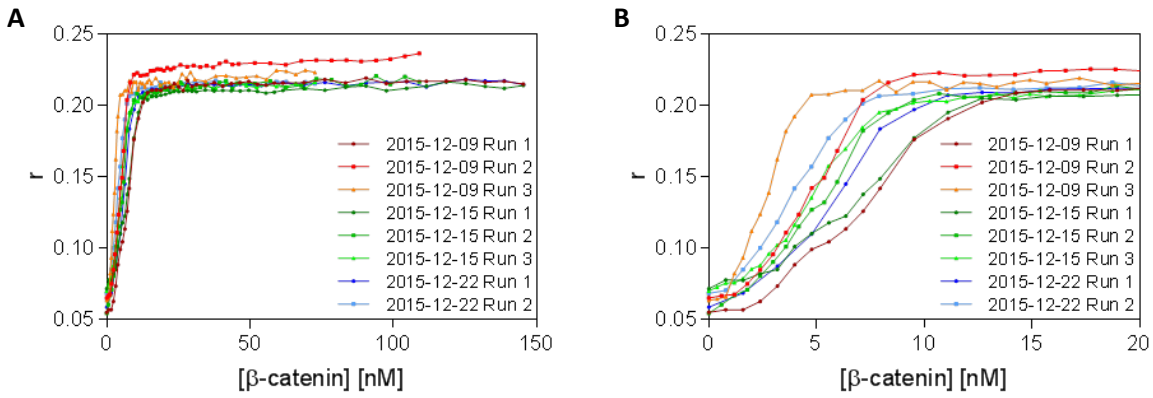


Figure 4-6 – Fluorescence anisotropy titration measurements for "WT" flS31C binding to β-catenin. Eight independent traces of "WT" flS31C binding to β-catenin collected over three days. (B) shows the same data as (A) with the transitional part of the curves highlighted. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

4.6 Alternative Labelling Positions

When deciding where to introduce a cysteine in order to label with fluorescein maleimide two other serine residues were considered for SDM, S^{20} and S^{46}. Both of these residues form a single contact with β-catenin in all structures and so are less optimal than the S^{31} position, in terms of minimising steric hindrance from labelling, and also appear to be in near identical environments between the E^{24}-bound and E^{29}-bound structures. However, the main aim of this study is to elucidate a mode of binding for TCF7L2 and β-catenin, and for that it would be useful to have probes in different positions to report on local changes. The "WT" S46C and "WT" S20C constructs were made by SDM and their kinetic rate constants were determined in the same manner as the "WT" S31C construct (see Appendix A, Section A.3.2).

On testing, it was found that the "WT" flS46C had a slightly slower association rate, but also a slightly slower $k_{off,major}$ resulting in similar $K_{d,kin}$ values (summarised in Table 4-3, also see Figure 4-7a & b). For all three constructs the association traces were best described by a one-phase exponential

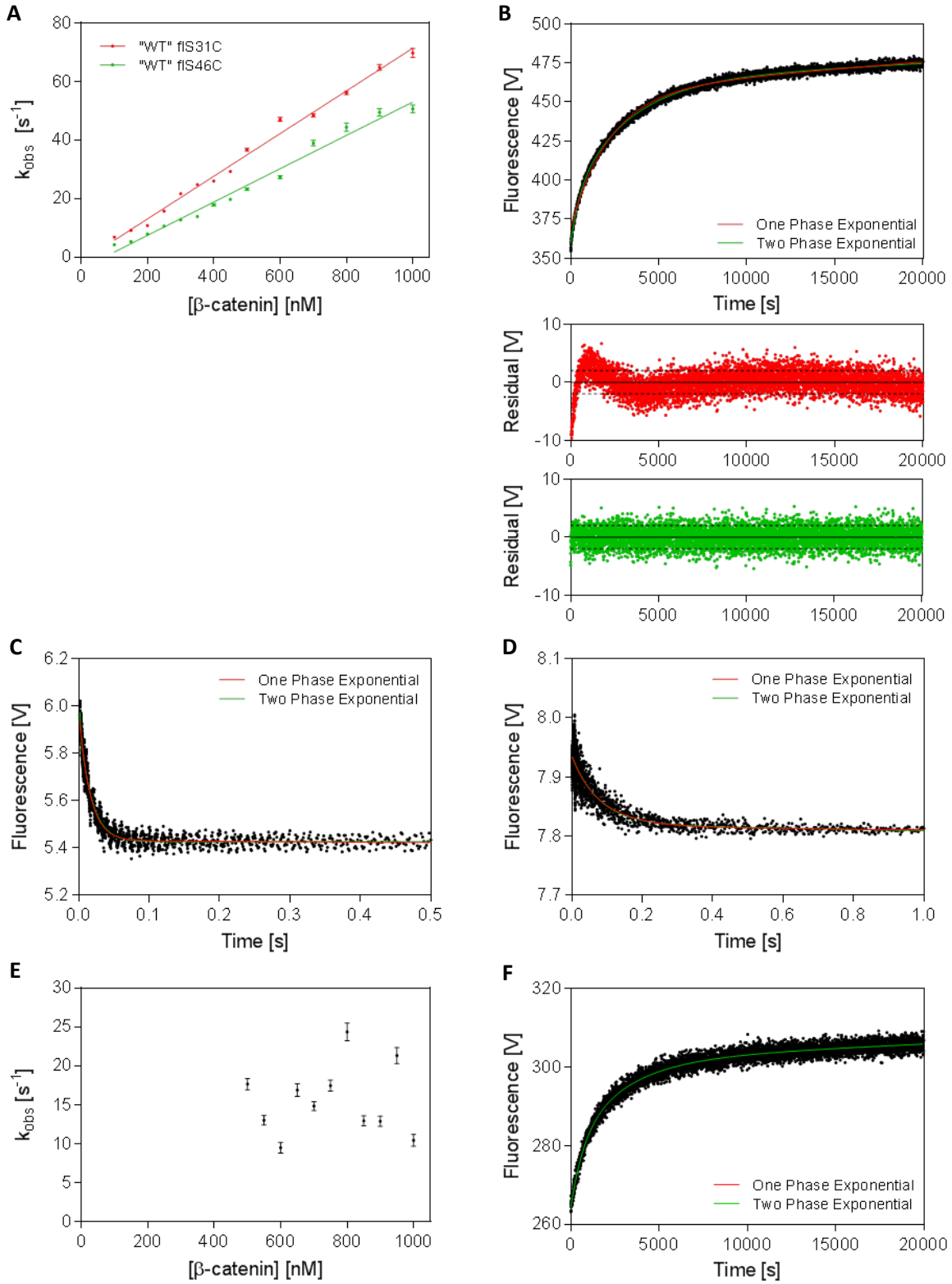


Figure 4-7 – The kinetic parameters of the "WT" fIS46C and "WT" fIS20C construct. (A) shows the association rate of "WT" fIS46C binding to β-catenin compared to "WT" fIS31C. (B) shows the dissociation rate of 200 nM "WT" fIS46C-β-catenin complex in the presence of 50 times molar excess of unlabelled competitor. (C) and (D) show the "WT" fIS31C and "WT" fIS20C constructs binding to β-catenin respectively at the same [β-catenin]. (E) shows a typical plot of k_{obs} against the concentration of β-catenin for the "WT" fIS20C construct, demonstrating no clear concentration dependence for the observed reaction. (F) shows a typical trace for the dissociation of the "WT"fIS20C-β-catenin complex. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

function and the dissociation traces by a two-phase exponential function. However for the "WT"

flS20C construct, the change in fluorescence intensity on association or dissociation was significantly lower than for either the S31C or the S46C construct, requiring that the PMT voltage be increased in order to see any signal, worsening the signal to noise ratio. Whereas the dissociation of the "WT" flS20C-β-catenin complex could be plotted and fit in the same manner as for the other two, this was not true for the association traces. When the k_{obs} for the association traces was plotted against the concentration of β-catenin to calculate the k_{on} in the usual manner, there appeared to be no concentration dependence to the observed rate constant meaning a k_{on}, and thus a $K_{d,kin}$, could not be calculated (see Figure 4-7c-f).

Table 4-3: Association and dissociation rate constants and $K_{d,kin}$ for "WT" flS46C and "WT" flS20C.

Construct	Association		Dissociation				$K_{d,kin}$ (pM)*
	k_{on} (x10^7 M^{-1} s^{-1})	n†	$k_{off,major}$ (x10^{-4} s^{-1})	$k_{off,minor}$ (x10^{-3} s^{-1})	% fast	n†	
"WT" S31C	7.33 ± 0.14	16	5.73 ± 0.40	1.52 ± 0.28	23 ± 3	8	7.8 ± 0.6
"WT" S46C	5.46 ± 0.12	4	3.17 ± 0.08	2.10 ± 0.05	26 ± 2	3	5.8 ± 0.4
"WT" S20C	n/a	4	4.74 ± 0.92	1.58 ± 0.48	37 ± 4	2	n/a

* $K_{d,kin} = k_{off,major}/k_{on}$
† n is the number of repeats

4.7 Long-Term Storage of Labelled TCF7L2

After several labelling reactions it became apparent that the amount of labelled TCF7L2 construct produced was more than required for most experiments. So excess labelled protein was flash frozen with liquid nitrogen and stored at -80°C for later use. However when the frozen labelled TCF7L2 constructs were thawed and used in experiments, it produced considerable variability in the results obtained (see Figure 4-8 and Table 4-4). We were unable to explain this occurance as later testing showed no obvious aggregation of the labelled protein, or change in excitation/emission or fluorescence intensity of the fluorescein mioety. Consequently for the rest of this project it was required that all TCF7L2 be freshly labelled before use.

4.8 Discussion

After this rigorous testing I felt confident in using the "WT" S31C construct as a pseudo wild-type baseline with which to compare other TCF7L2 constructs. In the ITC experiments I have shown that the introduction of the S31C mutation had had a negligible effect on the TCF7L2-β-catenin interaction, giving a K_d of 15 ± 4 nM compared with the 15 ± 6 nM and 16 ± 3 nM from the literature (Omer *et al.*, 1999; Sun and Weis, 2011). When this construct is labelled with fluorescein maleimide the K_d increases slightly to 28 nM, which is to be expected. The "WT" S31C construct is a small protein (5.9 kDa) with a high charge density (19 charged residues out of 56 total), and so conjugating this to a large (428 Da), hydrophobic fluorophore is likely to affect its dynamic properties.

Table 4-4: Association constants for fresh and frozen, "WT"S31C and "WT" S46C constructs in PBS buffer.

"WT" S31C		"WT" S46C	
A	**k_{on} (x10^7 M^{-1} s^{-1})**	**B**	**k_{on} (x10^7 M^{-1} s^{-1})**
Fresh	3.69 ± 0.21	Fresh	4.19 ± 0.17
50 days frozen	5.49 ± 0.17	5 days frozen	5.12 ± 0.24
84 days frozen	6.39 ± 0.46	9 days frozen	1.70 ± 0.04
111 days frozen	7.24 ± 0.18	13 days frozen	3.06 ± 0.15
C	**k_{on} (x10^7 M^{-1} s^{-1})**	**D**	**k_{on} (x10^7 M^{-1} s^{-1})**
2015-12-09	6.76 ± 0.40	2015-12-09	5.69 ± 0.21
2015-12-15	8.23 ± 0.31	2015-12-15	6.06 ± 0.17
2015-12-22	7.29 ± 0.19	2015-12-22	5.76 ± 0.33
Weighted Average	7.33 ± 0.14	Weighted Average	5.46 ± 0.12

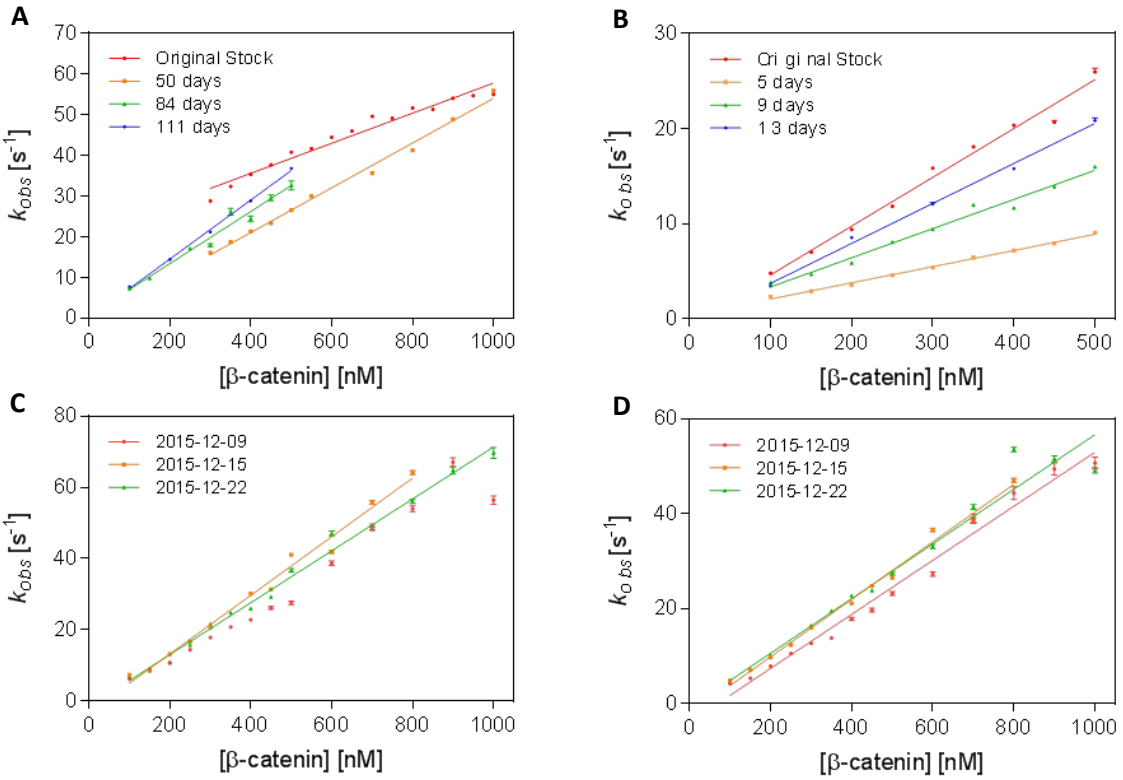


Figure 4.8 – The effect of freezing on labelled TCF7L2 constructs binding to β-catenin. (A) and (B) show how the association rate (k_{on}) of the same batch of fluorescein maleimide labelled TCF7L2 binding to β-catenin is affected by freezing for different lengths of time for the "WT" S31C construct and the "WT" S46C construct respectively. (C) and (D) compare how association rate varies different batches of freshly labelled TCF7L2 constructs binding to the same batch of β-catenin for the "WT" S31C construct and the "WT" S46C construct respectively. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

When considering which fluorophore to use, a number were considered but ultimately fluorescein seemed optimal. With the added inefficiency of having to freshly label TCF7L2 before every experiment, price was a major factor in this decision. While there are certainly fluorophore which

are smaller, and so are less likely to have an effect on binding dynamics (e.g. coumarin-derived fluors), or which are less hydrophobic, and so are less likely to affect solubility (e.g. Alexa fluors), these factors did not seem to have a significant effect on monitoring the TCF7L2-β-catenin interaction. Fluorescein maleimide was cheap, had convenient excitation and emission wavelengths and the protocols for labelling and the removal of excess dye are quick and simple.

What is more concerning is the difference between the K_d measured at equilibrium using the ITC and the kinetic K_d, $K_{d,kin}$, obtained from the association and dissociation kinetics. For the "WT" flS31C in SunWeis buffer, the $K_{d,kin}$ was 8.3 ± 0.6 pM. While in this case the $K_{d,kin}$ will be less certain, due to the dissociation being multi-phasic, this value is still three orders of magnitude tighter binding than that obtained by ITC. Even when the kinetic K_d is calculated from the faster minor phase, this gives a K_d of 19.5 ± 2.3 pM, still three orders of magnitude different for the ITC K_d.

Although unusual, there have been previous reports of naturally occurring binding interactions as strong as this for both globular proteins (streptavidin-biotin, K_d ≈ 1 fM (Weber *et al.*, 1989), zinc-hook proteins, K_d ≈ 1 fM (Kochańczyk, Jakimowicz and Krężel, 2013)) and intrinsically disordered proteins (S-peptide-S-protein, K_d ≈ 6 pM (Goldberg and Baldwin, 1998), PUMA BH3-Mcl1 , K_d ≈ 100 pM (Rogers, Steward and Clarke, 2013)). Similarly, IDPs/IDRs with association rate constants similar to that of TCF7L2 (ACTR-NCBD, k_{on} ≈ 5.9 x10^7 M^{-1} s^{-1} (Dogan *et al.*, 2012)) and even 1-2 orders of magnitude higher (HIF-TAZ1, k_{on} ≈ 1.3 x10^9 M^{-1} s^{-1} (Sugase *et al.*, 2007), AD2 1-TAZ2, k_{on} ≈ 1.7 x10^{10} M^{-1} s^{-1} (Arai, Ferreon and Wright, 2012)) have been reported as well as interactions with similar dissociation rate constants (S-peptide-S-protein, k_{off} ≈ 1 x10^{-4} s^{-1} (Goldberg and Baldwin, 1998), PUMA BH3-Mcl1 , k_{off} ≈ 1.6 x10^{-3} s^{-1} (Rogers, Steward and Clarke, 2013))

I believe the source of this discrepancy lies in the exceptionally long dissociation times. The major and minor dissociation rate constants for "WT"S31C in SunWeis buffer were 5.73 ± 0.12 and 11.9 ± 0.11 x10^{-4} s^{-1} respectively, which correspond to half-lives of 1370 ± 30 s (22.8 min) and 582 ± 56 s (9.7 min) respectively. These may seem excessively long for a biological system but can be rationalised in context. Transcription factors typically form large multi-component complexes, which logically would require a stable, long-lived core around which to assemble correctly for proper function. The TCF7L2-β-catenin interaction is the first step in the formation of a transcription factor complex and so it makes sense that it is stable (K_d ≈ 8 pM) and long-lived (dissociation half-life ≈ 23 min).

It was unfortunate that the fluorescence anisotropy titration experiments proved to be unreliable and failed to corroborate either the ITC or kinetic experiments. However on reflection, I think this

may have been due to problems with experimental design. When designing an anisotropy experiment, the concentration of one of the reactants should be kept constant at a concentration just below the K_d (Rossi and Taylor, 2011). In my experiment labelled TCF7L2 construct was kept at a constant concentration of 10 nM, which would have been acceptable had the K_d turned out to be in the region of 28 nM as suggested by ITC. On the contrary, had the K_d been in the region of 8 pM as suggested by kinetics this would have been inappropriate, but also the concentration of TCF7L2 that would have been suitable (e.g. 2 pM) would likely not produce a large enough signal to be measurable.

It is also generally advised when designing equilibrium experiments like ITC to allow the reaction mixture to incubate for five times the longest kinetic half-life to allow the system to actually reach equilibrium, which for this system means approximately 2 hr. For a plate-based assay this is easy to do, but for titration experiments like ITC and fluorescence anisotropy titration it is impractical. I allowed 500 s between injections in my ITC experiments and only 90 s in my fluorescence anisotropy titrations which by this criterion is not long enough. I had considered a plate-based approach to measure the fluorescence anisotropy, but decided against it for the same reason I rejected ITC.

The similarity between the results obtained for different buffers was reassuring. Any differences can be explained by the difference in salt concentration (100 mM NaCl for SunWeis, 140 mM NaCl for PBS), which is known to affect protein-protein interactions by disrupting electrostatic interactions (Sheinerman, Norel and Honig, 2000). PBS is a more universal buffer system than that used by Sun and Weis (25 mM Tris pH 8.8, 100 mM NaCl, 2 mM DTT) or Omer *et al.* (50 mM HEPES pH 7.5) which allows me to compare my findings not only to those of Drs. Rowling and Dunbar, but also to other protein-protein interactions detailed in the literature.

Going forwards I decided to continue using the S31C labelling position over S46C position as I believed it would be better at reporting the difference between the E^{24}-bound and E^{29}-bound TCF7L2-β-catenin complexes, although it was never ruled out for further experiments. For many of the TCF7L2 variants detailed in later chapters, both to S31C and S46C variants were made but only the S31C variants were transformed, expressed and purified.

The S20C variant is still interesting. The concentration independence of the association rate indicates that it is not the same TCF7L2-β-catenin binding event that is being detected by the label at this site, but rather a structural rearrangement after association. However it is also possible that the much poorer signal-to-noise ratio means that the kinetic traces cannot be accurately plotted. Indeed, when the data for all four repeats of the “WT” flS20C association experiments are plotted on the

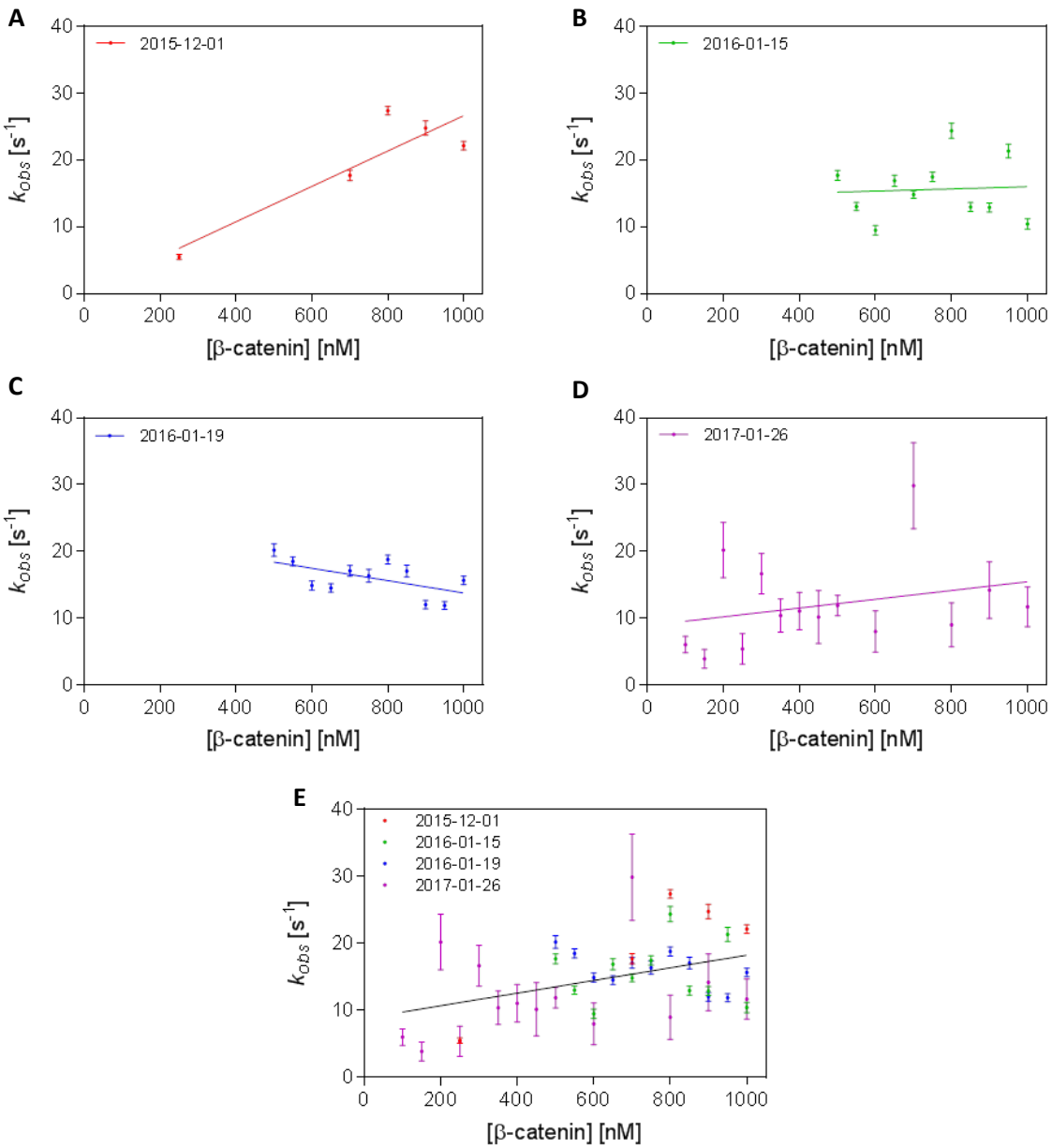


Figure 4-9 – All association data for the "WT" fIS20C construct. (A), (B), (C) and (D) are the four separate association experiments for "WT" S20C construct associating to β-catenin. When all four are plotted on the same axes (E) there appears to be a weak [β-catenin] dependence, but when compared to the individual traces that this is not the case. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

same graph then there is a hint that there is a [β-catenin] dependence on the observed association rates (see Figure 4-9e). The best fit for these amalgamated data gives a k_{on} of 0.94 ± 0.34 x10^{7} M^{-1} s^{-1} compared to 7.33 ± 0.14 x10^{7} M^{-1} s^{-1} for the S31C construct. However the R^2-value for this fit is only 0.165, and when compared to the individual traces (see Figures 4-9a-d), it is apparent that there is no [β-catenin] dependence.

Chapter 5 – Three Crystal Structures and Ensemble Binding

In the PDB there exist three unique crystal structures of TCF7L2 bound to β-catenin which have noteworthy differences between them. Two of these structures, those obtained by Poy *et al.* (PDB ID 1JPW) and Graham *et al.* (PDB ID 1JDH), are the most dissimilar and curiously were simultaneously reported in the same issue of Nature Structural Biology (Graham *et al.*, 2001; Poy *et al.*, 2001). A later crystal structure of a ternary complex comprising TCF7L2, β-catenin and the co-activator BCL9 (PDB ID 2GL7, see Section 1.5.2) was obtained by Sampietro *et al.* (Sampietro *et al.*, 2006).

Sampietro *et al.*'s structure is most similar to that obtained by Poy *et al.*'s, both of which show the glutamate at position 24 of TCF7L2 interacting with the lysine at position 312 of β-catenin. However in Graham *et al.*'s structure it is the glutamate at position 29 of TCF7L2 that interacts with lysine-312. This is due to residues 21-32 forming an α-helix in Graham's crystal structure, instead of extended structure, forcing glutamate-24 to point away from the surface of β-catenin and positioning glutamate-29 in the equivalent position where it can interact with lysine-312 (see Figure 5-1). Previous studies of β-catenin in complex with its binding partners have identified lysine-312 and lysine-435 as important (sometimes referred to as the "lysine buttons" or "charged buttons") so it seems unusual that there should be such variation around this residue (Graham *et al.*, 2000; Dyson and Wright, 2002; Gail, Frank and Wittinghofer, 2005; Xu and Kimelman, 2007; Sun and Weis, 2011).

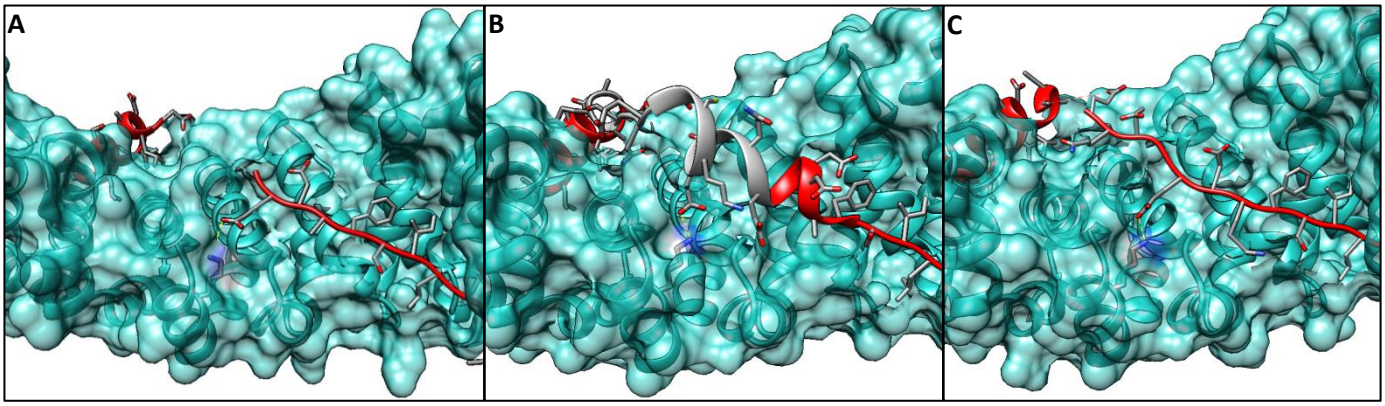


Figure 5-1 – The differences in crystal structures of TCF7L2 binding to β-catenin. (A), (B) and (C) are the crystal structures of TCF7L2 (red) in complex with β-catenin (cyan) obtained by Poy *et al.*, Graham *et al.* and Sampietro *et al.* respectively. β-catenin K^{312} and its interaction with the glutamates of TCF7L2 are highlighted. In structure (B), residues which are unstructured and unseen in (A) are coloured grey. Structure images (A) (B) and (C) were created using UCSF Chimera (Pettersen *et al.*, 2004) from PDB IDs 1JPW (Poy *et al.*, 2001), 1JDH (Graham *et al.*, 2001) and 2GL7 (Sampietro *et al.*, 2006) respectively. (D) is a map of how major structural elements differ between the three crystal structures, by residue.

In Graham's structure we can see the position of every TCF7L2 residue between 12 and 49, whereas in the other two structures a significant number of residues in this region are disordered and not seen in the crystal structure (residues 27 to 39 in Poy's structure and residues 29 to 37 in Sampietro's structure). Again this is due to the formation of a second α-helix. Additionally the conserved C-terminal α-helix (residues 41-49) has an extra turn (residues 37-40), taken together these create the effect of TCF7L2 being pulled taut like a string across the surface of β-catenin.

As was mentioned in Section 1.2.1 and in Chapter 4 alternative conformations of IDPs have the potential to affect IDP binding, e.g. p53, and the TCF7L2-β-catenin complex is known to recruit a range of other proteins. We speculate that these alternative structures present a mechanism for distinguishing between binding partners.

In this Chapter, I investigate the differences between the E^{24}-interacting and E^{29}-interacting structures by measuring the effect of a series of mutational variants on the binding kinetics of the TCF7L2-β-catenin interaction using the methods developed in Chapter 4. I then investigate the impact of osmolality on the binding kinetics of the pseudo wild-type construct and mutational variant with a lower net charge.

5.1 E24A S31C and E29A S31C

In order to study the differences between the two crystal structures I decided to try to force the TCF7L2 constructs to preferentially adopt one structure over the other. To that end, I used SDM to create the E24A S31C and E29A S31C variants (see Appendix A, Section A.3.3). Without the glutamate at position 24, it was hoped that the E24A S31C construct would be more strongly biased towards adopting the E^{29}-bound structure, and *vice versa* for the E29A S31C construct. These constructs were labelled as described in Section 2.6.2, and their association and dissociation rate constants were determined as for the "WT" flS31C construct. For each it was found that the association traces could be fit to a single exponential function, and the dissociation traces were best fitted to a two-phase exponential function with a faster minor phase and a slower major phase, as was the case for the "WT" flS31C construct (see Figure 5-2, summarised in Table 5-1).

The E24A flS31C construct had very slightly slower association and dissociation rates compared with the "WT" S31C construct, resulting in a very similar $K_{d,kin}$ (see Table 5-1). The E29A flS31C construct had a similar association rate to the E24A flS31C construct but faster dissociation rates, resulting in a $K_{d,kin}$ roughly double that of the wild-type (see Table 5-1). This suggested that the E^{24}-bound and the E^{29}-bound structures are very similar in energies with the E^{24}-bound structure being slightly higher in energy and that TCF7L2 can easily switch between the two structures.

Table 5-1: Association and dissociation rate constants and $K_{d,kin}$ for E24A, E29A, and E24A E29A flS31C constructs.

Construct	Association		Dissociation				$K_{d,kin}$ (pM)*
	k_{on} (x10^7 M^{-1} s^{-1})	n†	$k_{off,major}$ (x10^{-4} s^{-1})	$k_{off,minor}$ (x10^{-3} s^{-1})	% fast	n†	
"WT"	7.33 ± 0.14	16	5.73 ± 0.40	1.52 ± 0.28	23 ± 3	8	7.8 ± 0.6
E24A	6.05 ± 0.12	6	5.03 ± 0.40	1.53 ± 0.21	26 ± 1	4	8.3 ± 0.7
E29A	5.73 ± 0.43	2	8.97 ± 0.44	2.45 ± 0.21	18 ± 2	5	15.6 ± 1.4
E24A E29A	6.13 ± 0.34	3	4.25 ± 0.29	1.73 ± 0.14	18 ± 4	3	6.9 ± 0.6

* $K_{d,kin} = k_{off,major}/k_{on}$
† n is the number of repeats

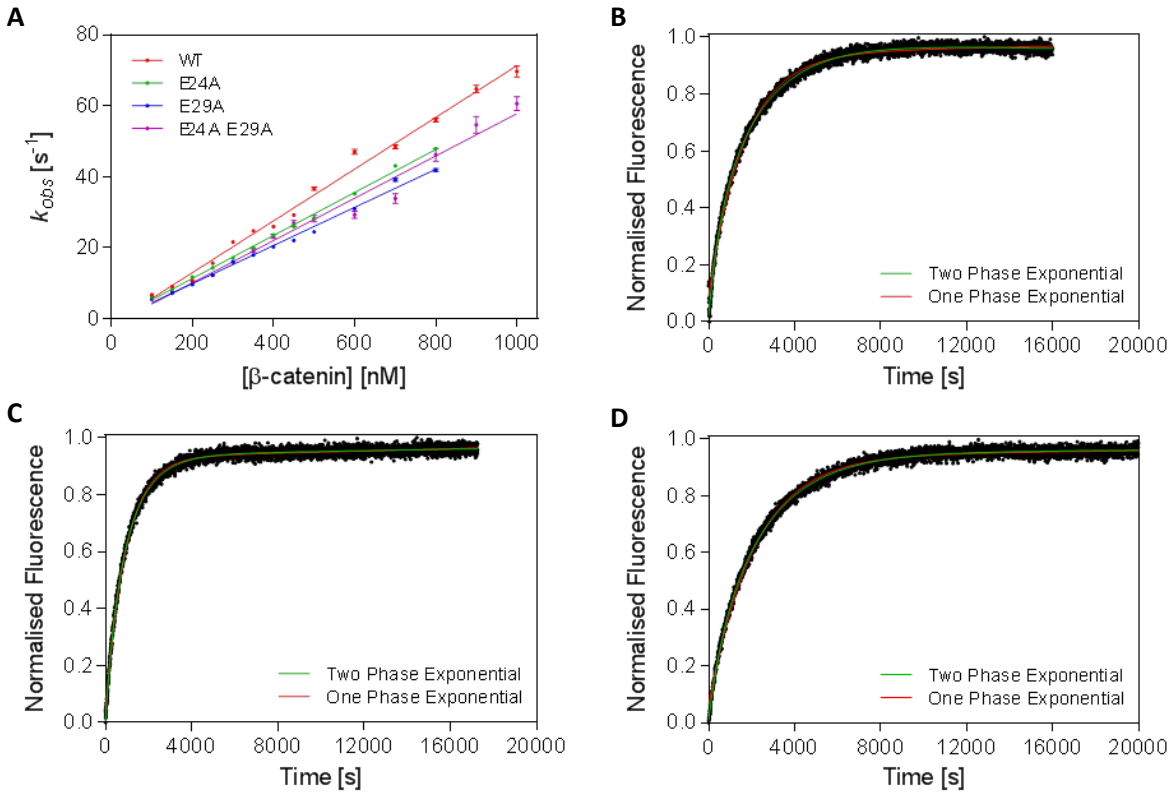


Figure 5-2 – The kinetic parameters of the E24A. E29A and E24A E29A constructs. (A) is a comparison of the association rate constants of three mutant constructs to the pseudo wild-type construct. (B), (C) and (D) are representative dissociation traces for the E24A, the E29A and the E24A E29A constructs respectively. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

5.2 E24A E29A S31C Double Mutational Variant

Assuming that the E^{24}-bound and the E^{29}-bound are the only structures that TCF7L2 can adopt when binding to β-catenin, then a E24A E29A S31C variant construct should not be able to form either structure and therefore should have a significant effect on the binding affinity. To test this hypothesis I made this E24A E29A S31C construct with SDM and tested it as I had done for the previous constructs. Once again it was found that association was one-phase and dissociation was two-phase.

Surprisingly the data showed that the E24A E29A double mutation had very similar kinetic rate constants to that of the E24A variant i.e. not significantly slower than those of the "WT" S31C construct, resulting in a similar $K_{d,kin}$ to the pseudo wild-type (see Figure 5-2, summarised in Table 5-1). This result is significant as it implies that TCF7L2 can adopt more structures when binding to β-catenin than the two captured in crystal structures.

5.3 Alternative Association Pathways

With the idea that other, undetected (*in crystallo*) structures could exist, I examined the sequence of TCF7L2 in greater detail and noticed that in the same region of glutamate-24 and glutamate-29 there are three other negatively charged residues that could theoretically interact with lysine-312 of β-catenin: aspartate-23, glutamate-26 and glutamate-28. E^{26} and E^{28} interacting with lysine-312 could easily be achieved through alternative positioning and length of the α-helix seen in Graham's structure alternatively presenting either one or the other of the two residues to the "lysine button". The D^{23}-bound structure would be more difficult to achieve as it would require further stretching out of the region N-terminal to it, which may weaken the affinity.

5.4 D23A S31C, E26A S31C, E28A S31C and XXXXX S31C

To investigate these potential alternative structures I produced and tested four more mutational variant constructs: D23A S31C, E26A S31C, E28A S31C and the quintuple mutational variant: D23A E24A E26A E28A E29A S31C, hereafter referred to as XXXXX S31C (see Appendix A, Section A.3.3). When tested all of these variants produced one-phase association traces and two-phase dissociation traces (see Figure 5-3, summarised in Table 5-2).

Table 5-2: Association and dissociation rate constants and $K_{d,kin}$ for the ensemble binding constructs.

Construct	Association		Dissociation				$K_{d,kin}$ (pM)*
	k_{on} (x10^7 M^{-1} s^{-1})	n†	$k_{off,major}$ (x10^{-4} s^{-1})	$k_{off,minor}$ (x10^{-3} s^{-1})	% fast	n†	
"WT"	7.33 ± 0.14	16	5.73 ± 0.40	1.52 ± 0.28	23 ± 3	8	7.8 ± 0.6
E24A	6.05 ± 0.12	6	5.03 ± 0.40	1.53 ± 0.21	26 ± 1	4	8.3 ± 0.7
E29A	5.73 ± 0.43	2	8.97 ± 0.44	2.45 ± 0.21	18 ± 2	5	15.6 ± 1.4
E24A E29A	6.13 ± 0.34	3	4.25 ± 0.29	1.73 ± 0.14	18 ± 4	3	6.9 ± 0.6
D23A	5.31 ± 0.25	3	3.00 ± 0.16	2.39 ± 0.70	15 ± 1	3	5.7 ± 0.4
E26A	5.59 ± 0.22	3	4.74 ± 0.30	1.66 ± 0.16	24 ± 4	3	8.5 ± 0.6
E28A	5.84 ± 0.35	4	4.96 ± 0.35	1.16 ± 0.28	23 ± 5	3	8.5 ± 0.8
XXXXX	4.26 ± 0.17	4	2.95 ± 0.23	2.34 ± 0.33	15 ± 2	4	6.9 ± 0.6

* $K_{d,kin} = k_{off,major}/k_{on}$
† n is the number of repeats

Unexpectedly, I found that none of these constructs, even the XXXXX S31C construct, had significantly different rate constants from that of the pseudo wild-type construct. All of these ensemble binding constructs had slightly slower association rate constants than the pseudo wild-type, with the XXXXX S31C construct being the slowest at approximately half that of the pseudo wild-type. All of these constructs, with the exception of the E29A S31C construct, also had slightly slower $k_{off,major}$ than the pseudo wild-type, again with the XXXXX S31C construct being the slowest at approximately half that of the pseudo wild-type.

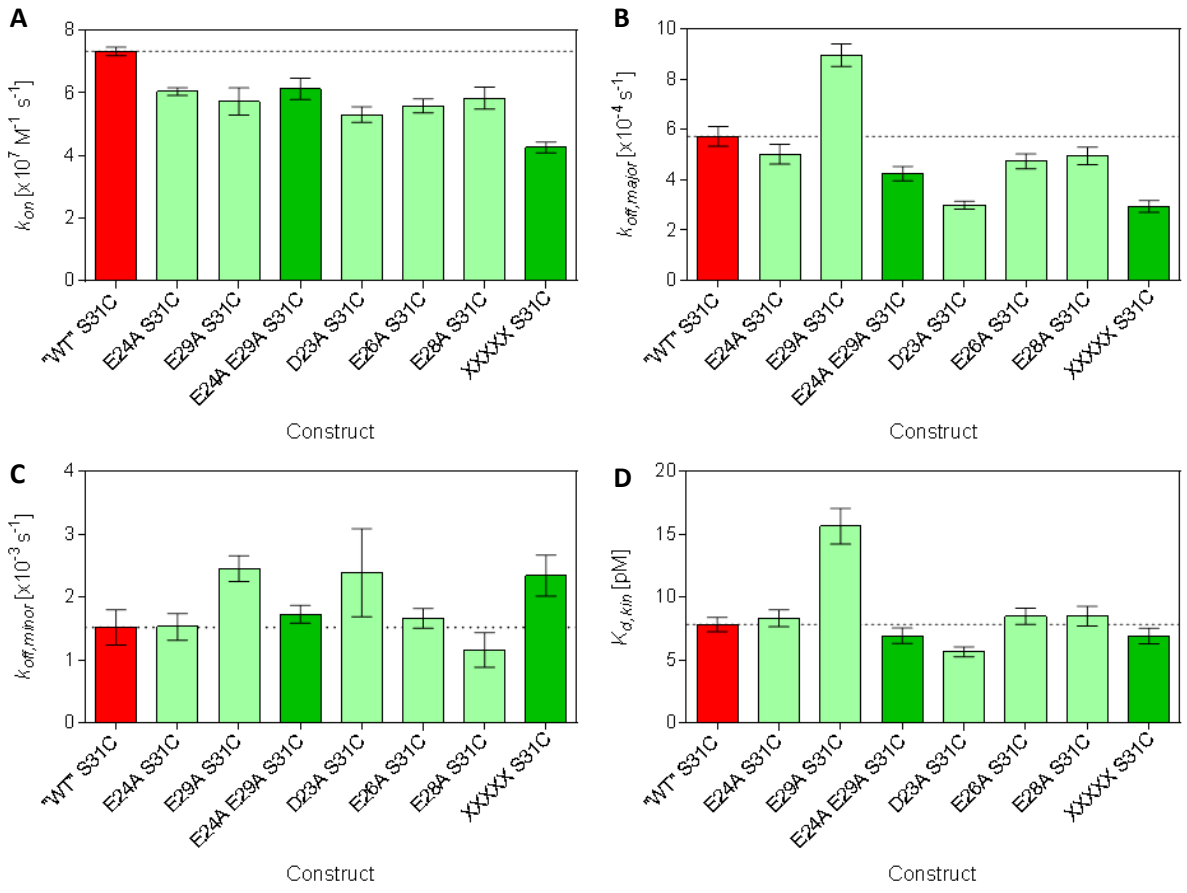


Figure 5-3 – The kinetic parameters of the ensemble binding constructs. (A), (B), (C) and (D) are comparisons of the association rate constants, major and minor dissociation rate constants and kinetic K_d respectively for all the ensemble binding mutant constructs. All experiments were performed in 1x PBS + 1 mM DTT.

5.5 The Effect of Charge on the TCF7L2-β-Catenin Interaction

It seems very unusual that the XXXXX S31C variant, in which five negative charges have been removed, should have such a negligible effect on the binding interaction given that electrostatic interactions are one of the major contributors to protein-protein binding energy. Electrostatic forces tend to have a greater influence on protein-ligand interactions (including protein-protein) than they do in protein folding (Sheinerman, Norel and Honig, 2000) because when folded, cytosolic proteins bury the majority of their hydrophobic residues and present mostly hydrophilic and charged residues

on their surface, in order to remain soluble. IDRs/IDPs are not folded by nature and tend to be enriched in charged residues, as they are typically disorder-promoting as well as increasing solubility (Dunker *et al.*, 2001) and so one would expect IDP-partner interactions to have a significant electrostatic component to them.

A way of determining the extent to which electrostatic interactions influence protein-protein binding is to vary the ionic strength of the buffer, usually by changing [NaCl], and measuring the changes in the K_d, and the association and dissociation rate constants. The magnitude of these variations can be determined by the slopes of –log(K_d) (or log(K_a)), log(k_{on}) and log(k_{off}) verses log([NaCl]), the gradients for which are referred to as Γ^{Eq}_{NaCl}, Γ^{on}_{NaCl}, and Γ^{off}_{NaCl} respectively (Vindigni *et al.*, 1997; Grucza *et al.*, 2000; Chemes, Sánchez and de Prat-Gay, 2011). A negative value of Γ^{Eq}_{NaCl} and Γ^{on}_{NaCl}, and a positive value for Γ^{off}_{NaCl} indicate that the complex is destabilised by increasing [NaCl].

$$\Gamma^{Eq}_{NaCl} = -\frac{\partial \log(K_d)}{\partial \log([NaCl])} = \frac{\partial \log(K_a)}{\partial \log([NaCl])} \quad [5\text{-}1]$$

$$\Gamma^{on}_{NaCl} = \frac{\partial \log(k_{on})}{\partial \log([NaCl])} \quad [5\text{-}2]$$

$$\Gamma^{off}_{NaCl} = \frac{\partial \log(k_{off})}{\partial \log([NaCl])} \quad [5\text{-}3]$$

As Γ^{on}_{NaCl} is a measure of how much the transition state is affected by variations in [NaCl], and Γ^{Eq}_{NaCl} is a measure of how the complex stability is affected by [NaCl], we can use these values to do an analogous analysis to Φ-value analysis (discussed in much greater detail in Section 6.3), to measure what proportion of the energy in the rate-limiting transition state comes from electrostatic interactions, α_{NaCl} (Chemes, Sánchez and de Prat-Gay, 2011).

$$\alpha_{NaCl} = \frac{\partial \log(k_{on})/\partial \log([NaCl])}{-\partial \log(K_d)/\partial \log([NaCl])} = \frac{\Gamma^{on}_{NaCl}}{\Gamma^{Eq}_{NaCl}} \quad [5\text{-}4]$$

I used this method to test both my “WT” S31C construct and my XXXXX S31C construct, as I believed that the removal of five negatively charged residues would produce a significant change in these values from the pseudo wild-type. However, as with the Φ-value analysis (see Section 6.3), the values obtained for α_{NaCl} are qualitative given I was unable to determine the equilibrium K_d for my TCF7L2 constructs binding to β-catenin.

Buffers of the required salinity were prepared by mixing 10x PBS stock, 5 M NaCl stock and 1 M DTT to give final concentrations 1x PBS pH 7.4, 1 mM DTT + additional NaCl. These calculations were made under the assumption that 1x PBS already is 140 mM NaCl, so for example to make a solution

that had a final [NaCl] of 1 M, I added and amount of 5 M NaCl stock that would provide and addition 860 mM NaCl.

For the association traces, lyophilised β-catenin and labelled TCF7L2 construct were both resuspended in the buffer with the desired final concentration of NaCl as described in Section 2.7, diluted to the required concentrations with the same buffer and then mixed in 1:1 ratios as described in Section 2.8. Association reactions were performed as described in Section 2.8, and the association traces for both constructs fit best to a one-phase exponential function at all combinations of [β-catenin] and [NaCl].

For dissociation traces this mixing method was not possible, as the concentrated, unlabelled wild-type TCF7L2 (1-54) that I used as a competitor was stored in 1x PBS, 1 mM DTT and trying to buffer exchange it into high salt PBS would result in an unknown final concentration of competitor. So instead β-catenin and labelled TCF7L2 construct were prepared in a buffer with a higher [NaCl] such that, when diluted 1:1 with the competitor in 1x PBS, 1 mM DTT (final concentration ≈ 140 mM NaCl), would have the desired final [NaCl]. The TCF7L2 and β-catenin were pre-complexed in the higher salt buffer before being mixed 1:1 with the competitor as described in Sections 2.8 and 2.9. Dissociations reactions were performed as described in Sections 2.8 and 2.9, as it was found that for both constructs at the highest [NaCl] tested (800 mM final [NaCl]), the stopped flow was required to measure the dissociation rate.

From these salt dependence experiments we can see that association rate decreased with increasing [NaCl] for both constructs and that the dissociation rate increased with increasing [NaCl], causing a rapid decrease in $K_{d,kin}$, i.e. complex stability, with increasing [NaCl] (see Figure 7-1). Also the magnitude of the minor fast phase decreases with increasing [NaCl]. However for each of the Γ-values, the pseudo wild-type construct was affected to a greater degree than the XXXXX variant, and that the XXXXX mutation had a much larger impact on the salt dependence of dissociation than on the salt dependence of association.

Table 5-3: [NaCl] dependence of "WT" S31C and XXXXX S31C constructs.

Construct	[NaCl] Dependence				
	$\Gamma^{on}{}_{NaCl}$	$\Gamma^{off,major}{}_{NaCl}$	$\Gamma^{off,minor}{}_{NaCl}$	$\Gamma^{Kd,kin}{}_{NaCl}$*	$\alpha_{NaCl}{}^{\dagger}$
"WT"	-1.72 ± 0.08	1.50 ± 0.13	1.89 ± 0.20	-3.22 ± 0.16	0.53 ± 0.04
XXXXX	-1.33 ± 0.07	0.61 ± 0.06	0.70 ± 0.12	-1.95 ± 0.14	0.68 ± 0.06

$*\Gamma_{NaCl}^{K_{d,kin}} = -\partial \log(K_{d,kin})/\partial \log([NaCl])$

$^{\dagger}\alpha_{NaCl} = \Gamma_{NaCl}^{on}/\Gamma_{NaCl}^{K_{d,kin}}$

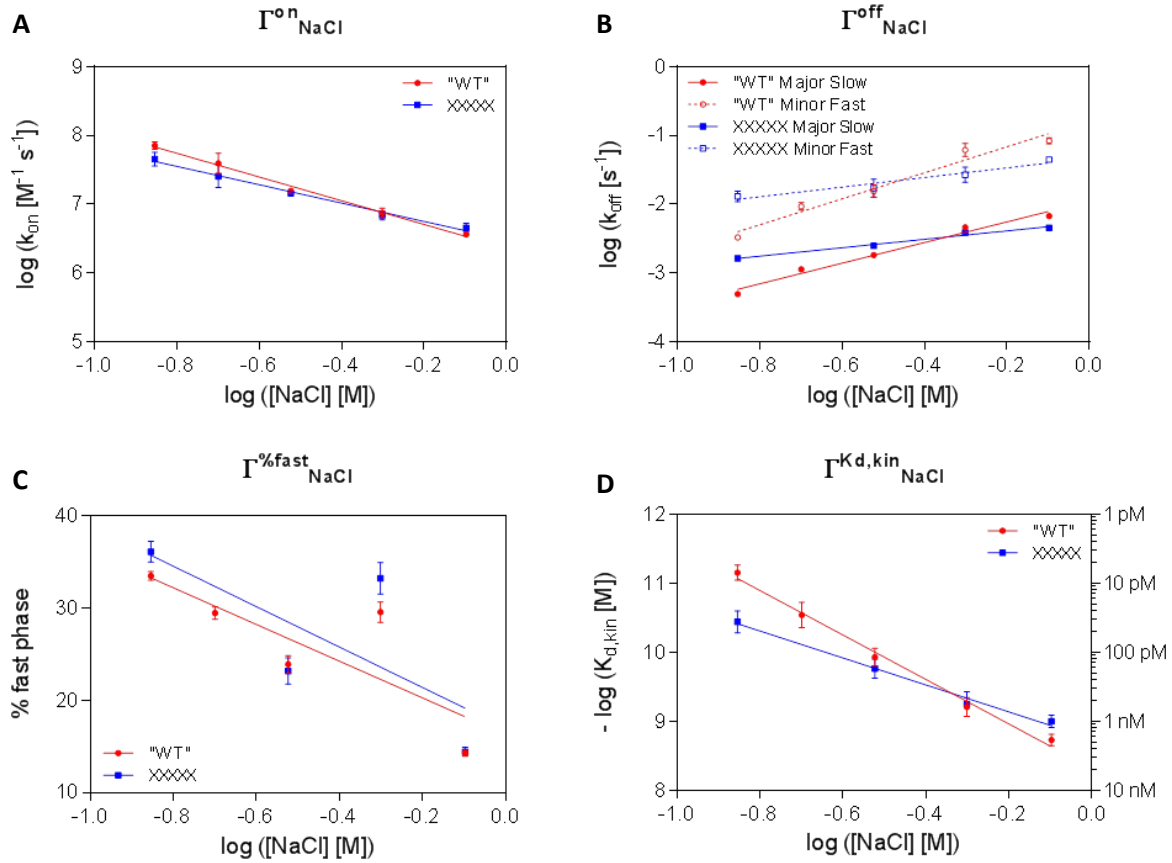


Figure 5-4 – The effect of [NaCl] on TCF7L2-β-catenin complex formation for "WT" S31C and XXXXX S31C constructs. (A), (B), (C) and (D) show the effects of increasing NaCl concentration on the association rate constant, the dissociation rate constants, the relative proportions of the major and minor dissociation rate constants and $K_{d,kin}$ respectively.

From the literature, the general trend for IDP-protein interactions is that increasing ionic strength decreases k_{on} whilst having minimal effect on k_{off} (Hemsath *et al.*, 2005; Chemes, Sánchez and de Prat-Gay, 2011; Japrung *et al.*, 2013; Rogers, Steward and Clarke, 2013; Dogan, Gianni and Jemth, 2014), with the S-peptide-S-protein interaction being a notable exception (Bachmann *et al.*, 2011, however Bachmann *et al.* speculate that this is an artefact caused by labelling with fluorescein, which results in a change in the net charge of the S-peptide from +1 to -1).

From my experiments, the XXXXX S31C construct fits this trend much more closely than my pseudo wild-type construct, for which increasing ionic strength does have a significant effect on both dissociation rate constants (see Table 5-3 and Figure 5-4b). This result suggests that that cluster of negatively charged residues from 23 to 29 can strongly influence the kinetics of TCF72-β-catenin binding, but at physiologically relevant ionic strengths they have little to no net effect on the interaction.

5.6 Discussion

Ensemble binding, or fuzzy complexes, is commonly observed mode of binding for IDPs and IDRs binding to a structured partner (Sharma *et al.*, 2015). IDPs have a lot of conformational flexibility and therefore forming a rigid structure upon binding to a partner incurs a very high entropic cost. By forming a more dynamic complex this cost is lessened whilst still providing a favourable enthalpic contribution to the binding interaction. However the simplest explanation of my observations is that this is not a case of ensemble binding.

Taken together, these data suggest that this region of TCF7L2 does not actually contribute towards the TCF7L2 binding interaction as none of the measured rate constants differ significantly from those of the pseudo wild-type construct. Moreover the fact that the association traces for these variants were single phase and the dissociations biphasic, like the pseudo wild-type, suggests that I was observing the same association and dissociation events.

These results lead me to conclude that the differences observed in the previously reported crystal structures (Graham *et al.*, 2001; Poy *et al.*, 2001; Sampietro *et al.*, 2006) are simply a result of the different conditions in which the protein was crystallised. While it may be true that these are real structures that TCF7L2 is capable of adopting when binding to β-catenin in solution, they appear to be transient and do not contribute to binding, and it is simply by chance that these are the three structures obtained crystallographically for this region of TCF7L2.

There are a range of methods that can be used to study the ensemble of structures adopted by IDPs, either as isolated proteins or in complexes, which could be applied to the TCF7L2- β-catenin complex (Peti *et al.*, 2018). Multiple techniques are frequently used in combination with each other, in order to obtain a more accurate picture of the possible structures. These include solution NMR and SAXS which are able to produce probability distributions for the positions of individual atoms, which can then be resolved into ensembles of statistically likely structures using computational techniques like EROS (ensemble refinement of SAXS). Other techniques like hydrogen-deuterium exchange mass spectroscopy (HDX-MS) are more useful for studying IDPs/IDRs in complex. HDX-MS uses MS to measure the rate at which backbone amide N–H groups exchange to N–D in a deuterated solvent (often D_2O). Areas where exchange is slow are areas that are less frequently solvent exposed so this technique can be used to measure what percentage of the time fuzzy binding regions are in contact.

It is interesting to note that my findings are in contrast to those of Graham *et al.*. In addition to determining the E^{29}-bound crystal structure (PDB ID 1JDH, Graham *et al.* (2001)), Graham *et al.* drew comparisons between their structure of human TCF7L2 in complex with human β-catenin and a

structure they had previously published, of *Xenopus* TCF7L1 in complex with *Xenopus* β-catenin (Graham *et al.*, 2000). As part of their comparative study, they performed mutagenesis and measured the effect on TCF7L2-β-catenin binding by immunoprecipitation. They found that the E24A variant reduced binding by 18-53%, the E29A variant reduced binding by 35-70%, the E24A E29A double variant by 85-90%, and a E24A E26A E28A E29A quadruple mutational variant (comparable to my XXXXX mutational variant though not equivalent) eliminated binding. However it should be said that immunoprecipitation is a qualitative method that relies on antibody binding to a specific epitope, and if that epitope becomes disrupted, in this instance if the epitope falls in the residues 23-29 region, then binding of the antibody will be lost.

It is also important note there was a flaw in my initial reasoning for this investigation. In order to determine the properties of any given structure you need to ensure that it adopts that structure and no other. In this case, where more than two possible structures can exist, the single alanine mutations do not force a specific structure they force "specifically not this structure". In hindsight, a better approach to examining the D^{23}-bound, E^{24}-bound, etc. structures specifically would have been to mutate to alanine the four other charged residues. Then I could be sure that the construct was indeed only capable of adopting the desired structure. Indeed I was in the process of creating the constructs by SDM (which I referred to as 1XXXX, X2XXX, XX3XX, XXX4X and XXXX5 respectively, 1XXXX being equivalent to the constructed used by Graham *et al.*) but decided to make and test the XXXXX construct in order to determine whether this avenue was worth pursuing. But as I observed no effect on binding for the XXXXX construct this line of investigation was discontinued.

Peptide stapling is another method I had considered as a way of forcing TCF7L2 to adopt the E^{29}-bound structure. By macrocyclising the TCF7L2 construct across the α-helix that is only present in the E^{29}-bound structure, then the conformational selection would be heavily biased towards the E^{29}-bound structure, effectively forcing it to adopt that structure. Experiments on peptide stapling are described in more detail in Section 7.1 of this Thesis.

The effect of charge on the TCF7L2-β-catenin interaction yielded some interesting results. A $\Gamma^{Kd,kin}{}_{NaCl}$ ≈ -3 for the pseudo wild-type construct means that the TCF7L2-β-catenin complex is destabilised by three orders of magnitude for every order of magnitude increase in ionic strength. This sensitivity is relatively high compared to other IDP-protein interactions ($\Gamma^{Kd,kin}{}_{NaCl}$ ≈ -1 for the ACTR-NCBD interaction (Japrung *et al.*, 2013), -1.4 > $\Gamma^{Kd,kin}{}_{NaCl}$ > -2.4 for various HPV E7 constructs binding to Rb (Chemes, Sánchez and de Prat-Gay, 2011)). In contrast it is only two orders of magnitude for the XXXXX S31C construct ($\Gamma^{Kd,kin}{}_{NaCl}$ ≈ -2) which this implies that the five negatively charged residues that have been removed in this construct are entirely responsible for this extra destabilising effect. This

change comes almost entirely from the unusual acceleration of disassociation ($\Gamma^{off,major}_{NaCl} \approx 1.5$ for pseudo-WT and $\Gamma^{off,major}_{NaCl} \approx 0.6$ for XXXXX S31C).

A possible insight into what might be happening here comes from studies on the ACTR-NCBD interaction. When ionic strength was increased, a charge reversal occurred in going from 0.02 M to 0.2 M (close to physiological) ionic strength, changing the overall charge of NCBD from positive to negative (Japrung *et al.*, 2013). The explanation given for this phenomenon was that the highly positively charged NCBD attracts a shell of negatively charge chloride ions. At higher ionic strengths sufficient chloride ions are attracted to this shell to completely screen and eventually overwhelm the positive charge. I believe an analogous behaviour occurs with TCF7L2-β-catenin, except instead we have a shell of positively charged sodium ions forming around a negatively charged region.

The decrease in Γ^{on}_{NaCl} upon mutation of the five charged residues is easy to explain as reducing the number of electrostatic interactions means that effects disrupting electrostatic interactions will have less of an impact. However the idea of an ionic shell/screen helps us explain the unusually high Γ^{off}_{NaCl}s for the pseudo wild-type construct, as well as the magnitude of the effect the XXXXX S31C mutation has on the Γ^{off}_{NaCl}s. As the variable region changes transiently between structures it would periodically part from the surface of β-catenin. While apart from β-catenin, the negatively charged residues can recruit a shell of positively charged sodium ions, which then inhibit re-association, promoting complete dissociation. This effect increases in magnitude with increasing ionic strength and is lost entirely when these residues are mutated to alanine in the XXXXX S31C construct.

The change in α_{NaCl} values between constructs can also be explained by the ionic shell model. Upon removing five charged residues (XXXXX S31C), α_{NaCl} increased, implying that the energetic contribution to binding from electrostatic interactions increases. As this cannot be reasoned as being an increase in stabilising interactions, it must be due to a decrease in disruptive interactions, a phenomenon which is seen again in Chapter 6. As we have already established that β-catenin is primarily positively charged and the removed TCF7L2 residues are negative, this should result in a decrease of stabilising interactions, unless, of course, they recruit this positively charged shell. It could also be the case that it is the other way around, and it is β-catenin that recruits a shell of negatively charge chloride ions instead, but this is less likely simply due to the difference in charge density, the 23-29 region of TCF7L2 being much more charge dense than the surface of β-catenin.

While the concept of a hydration/solvation shell is not new in protein science (Burling *et al.*, 1996), Japrung *et al.*'s idea of an ion shell affecting protein kinetics is somewhat unusual but I believe fits what is being observed here. In their paper describing the ACTR-NCBD interaction, Japrung *et al.*

performed single molecule studies on the system by attracting single NCBD molecules into nanopores using and electric current then adding ACTR (Japrung *et al.*, 2013). However they observed that at higher KCl concentrations, the NCBD would not be drawn into the pore with the current they had been using and they needed to reverse the current in order achieve the same effect. Potentially a similar experiment could be performed on a fragment of TCF7L2 containing the highly charged region to see if significant electrostatic shielding can be observed.

Chapter 6 – Alanine Scan and Φ-value Analysis

In Chapter 5 I had shown that, somewhat surprisingly, the highly negatively charged region of TCF7L2 (residues 23-29) has no significant effect on the TCF7L2-β-catenin binding interaction. This is rather unexpected as studies of β-catenin binding to its other intrinsically disordered partners show that binding to the β-catenin "lysine button", lysine-312, is important to their respective interactions (Graham *et al.*, 2000; Dyson and Wright, 2002; Gail, Frank and Wittinghofer, 2005; Xu and Kimelman, 2007; Sun and Weis, 2011). Yet TCF7L2 still forms a very tight interaction with β-catenin, the binding energy for which must therefore be provided by other residues of TCF7L2. Consequently, I chose to investigate the role of the other residues play in this binding interaction.

The TCF7L2-β-catenin binding interface is very long, with many different TCF7L2 residues forming contacts with a large number of different β-catenin residues, which I divided into three distinct binding regions; the N-terminal region (residues 12-21), the variable region (residues 22-37), and the C-terminal region (residues 38-50). While each residue will likely provide a contribution to the binding, it is likely that the majority of the binding energy comes from a few "hot-spot" residues (Clackson and Wells, 1995; Bogan and Thorn, 1998).

Indeed, Sun and Weis determined a consensus "lysine button" binding motif for β-catenin's disordered partners (Sun and Weis, 2011; Reichen, Hansen and Plückthun, 2014). That sequence is $Dx\Phi\Phi x\Omega x_{2\text{-}7}E$, where the initial aspartate residue interacts with "lysine button" lysine-435, Φ is a hydrophobic residue, Ω is an aromatic residue, and the terminating glutamate residue interacts with "lysine button" lysine-312. For TCF7L2, the initial aspartate residue corresponds to D^{16}, ΦΦ are L^{18} and I^{19} respectively, Φ is F^{21}, and the terminating glutamate is E^{24} or E^{29} depending on which crystal structure you're looking at.

In this chapter, I investigate the contribution of other regions of TCF7L2 to the TCF7L2-β-catenin binding interaction using a number of mutational variants as previously described in Chapter 5. These variants are then used to map the kinetic pathway of complex formation as well as investigate the potential for multiple association and dissociation pathways, analogous to the multiple folding and unfolding pathways observed for tandem repeat proteins.

6.1 Choice of Positions to Mutate to Alanine

In order to decide which residues to choose for an alanine scan I returned to the previously generated contact maps for the three crystal structures (see Figure 4-2). I looked for a mix of residues that formed the same or very similar contacts in each structure and some that were

different between structures. I also favoured residues that formed more contacts and had a larger contact areas as well as aiming to pick a mixture of charged, polar and hydrophobic residues. Another important consideration is that in order to perform Φ-value analysis, a mutation needs to significantly disrupt binding without eliminating it completely. For that reason I was hesitant about mutating residues within the consensus binding sequence described by Sun and Weis. In the end I decided on five residues from each of the N- and C-terminal binding regions: L12A, N15A, E17A, I19A and F21A, and E38A, L41A, V44A, K45A and L48A.

L^{12} and E^{38} were chosen as they both form multiple contacts in Graham's and Sampietro's structures but none in Poy's (see Figure 6-1), and seem to be delineating the beginning of their respective binding regions. This could be suggesting that these residues have more transient interactions and therefore provide a lesser contribution to overall binding, or it could be the effect of the different crystallisation conditions, as was likely the case for residues in the 23-29 region. The difference in

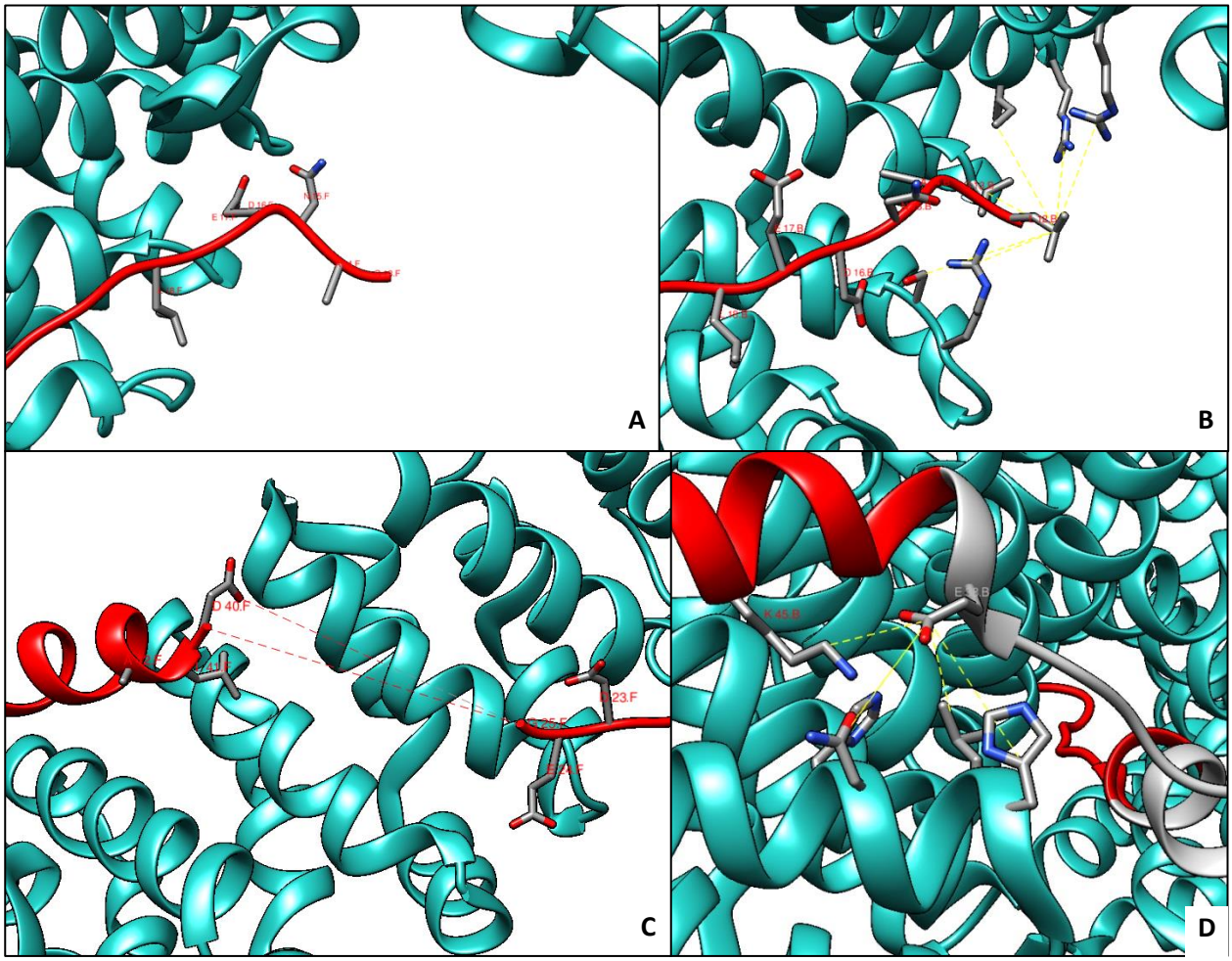


Figure 6-1 – TCF7L2 leucine-12 and glutamate-38 in crystal structures. (A) and (B) are zoomed in pictures of TCF7L2 (red) in complex with β-catenin (cyan) taken from Poy *et al.* and Sampietro *et al.* respectively. TCF7L2 residue L^{12} is not visible in Poy's structure but has been highlighted, along with its side chain interactions with β-catenin residues in Sampietro's structure. (C) and (D) are zoomed in pictures of TCF7L2 (red) in complex with β-catenin (cyan) taken from Poy *et al.* and Graham *et al.* respectively. Similarly, TCF7L2 residue E^{38} is not visible in Poy's structure but has been highlighted in Graham's. Structure images were created using UCSF Chimera from PDB IDs 1JPW, 1JDH and 2GL7 respectively.

the number of contacts in the contact maps for these residues for Sampietro's and Graham's structure are due to variations in weaker contacts (>7 Å).

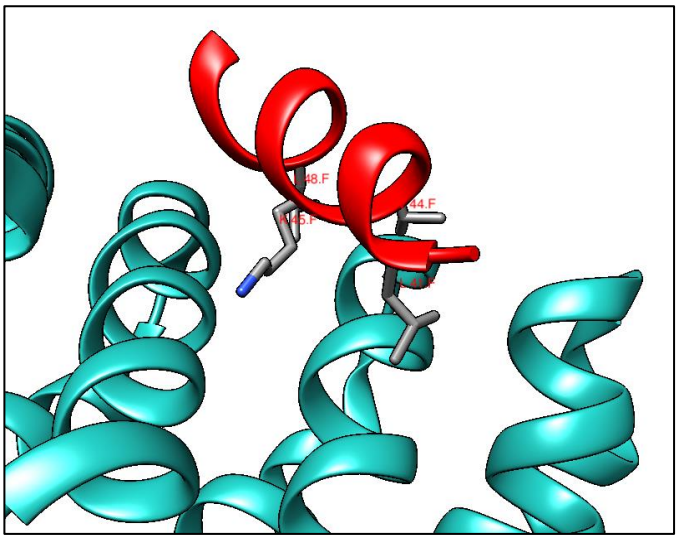


Figure 6-2 – TCF7L2 leucine-41, valine-44, lysine-45 and leucine-48 in crystal structure. A diagram highlighting the relative positions of residues L^{41}, V^{44}, K^{45} and L^{48} when TCF7L2 (red) binds to β-catenin (cyan). Structure images were created using UCSF Chimera from PDB IDs 1JPW (Poy *et al.*, 2001).

Residues L^{41}, V^{44}, K^{45} and L^{48} form the face of TCF7L2s C-terminal α-helix which contacts the surface of β-catenin and thus seemed the logical choices for an alanine scan. In all three crystal structures these four residues form contacts with same β-catenin residues suggesting that these interactions are important to TCF7L2-β-catenin complex formation (see Figure 6-2).

I purposely avoided the E16A mutation, as this residue contacts lysine-435 of β-catenin i.e. the other "lysine button". Previous studies of β-catenin binding to its other intrinsically disordered partners (Graham *et al.*, 2000; Dyson and Wright, 2002; Gail, Frank and Wittinghofer, 2005; Xu and Kimelman,

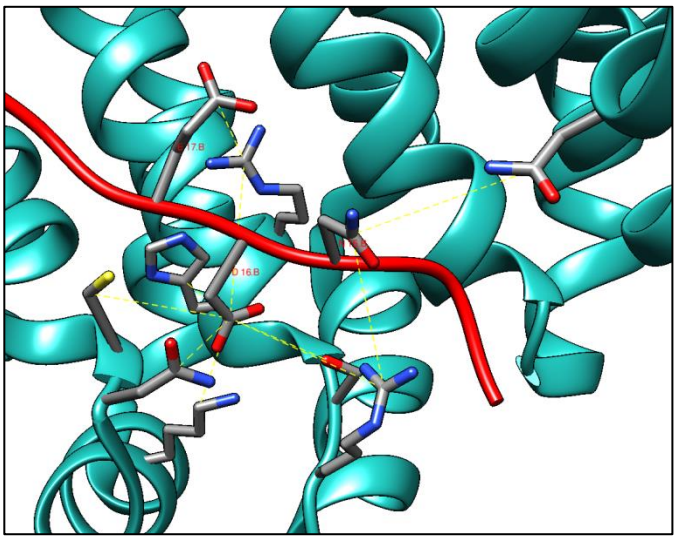


Figure 6-3 – TCF7L2 asparagine-15, apartate-16 and glutamate-17 in crystal structures. A zoomed in picture of TCF7L2 (red) in complex with β-catenin (cyan) taken from Poy *et al.*. The interactions of TCF7L2 residues N^{15}, D^{16} and E^{17} with β-catenin residues have been highlighted. Structure images were created using UCSF Chimera from PDB IDs 1JPW

2007; Sun and Weis, 2011) as well as Omer *et al.*'s original paper describing TCF7L2-β-catenin binding (Omer *et al.*, 1999), have identified this lysine button interaction as a hot spot. Mutating this residue may disrupt binding so much that it is no longer detectable, so I instead opted to mutate the adjacent residues N^{15} and E^{17}. From the crystal structures (see Figure 6-3) it can be seen that these two residues form well-defined interactions that should contribute significantly to binding.

I^{19} and F^{21}, in contrast, were chosen for their high contact area, hydrophobicity and their consistency between crystal structures (see Figure 6-4). F^{21} appears at first glance to be in different environments in the different crystal structures, being at the N-terminus of the transient α-helix in Graham's structure as opposed to the elongated structure seen in the other two. However, closer inspection shows that it is the aforementioned weaker contacts as well as some backbone interaction that vary between the structures.

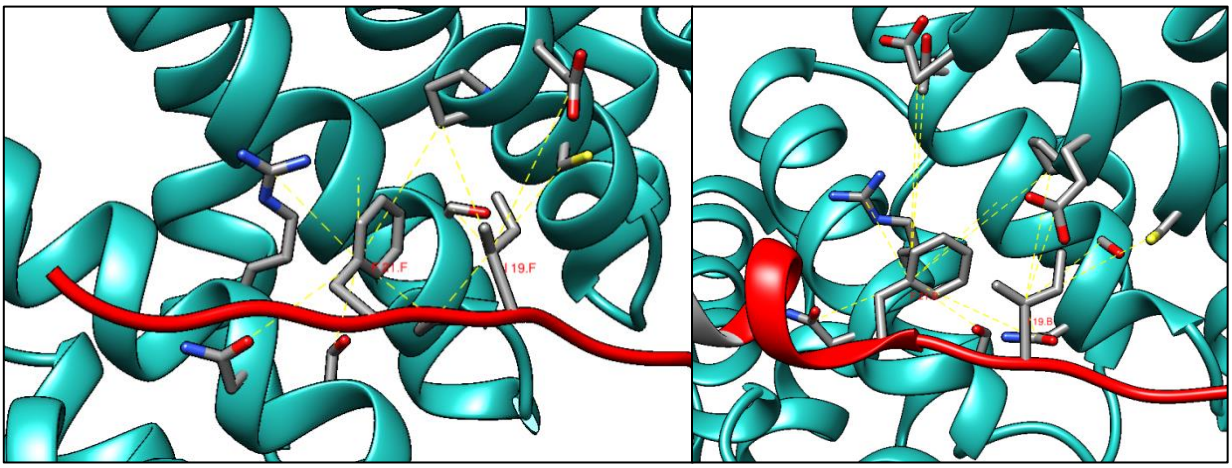


Figure 6-4 – TCF7L2 isoleucine-19 and phenylalanine-21 in crystal structures. (A) and (B) are zoomed in pictures of TCF7L2 (red) in complex with β-catenin (cyan) taken from Poy *et al.* and Graham *et al.* respectively, highlighting TCF7L2 I^{19} and F^{21} and their interactions with β-catenin residues. It can be seen that despite a difference in back bone conformation, the side chain of F^{21} occupies a similar position, maintaining the same side chain contacts. Structure images were created using UCSF Chimera from PDB IDs 1JPW and 1JDH respectively.

These ten mutations were introduced into the "WT" S31C construct by SDM (see Appendix A, Section A.3.4).

6.2 Alanine Scan

The association rate constants for each of the ten alanine variants were determined as previously described, and once again it was found that the association traces were one-phase. Nine out of the ten mutations selected for this alanine scan showed decreases in association rate constants which were generally smaller in magnitude than the decreases observed for the variants used to test ensemble binding (Chapter 5), except those observed for the F21A S31C and V44A S31C constructs which were of comparable magnitude (see Figures 6-5a & 6-6a, summarised in Table 6-2). Worthy of

note is that the K45A variant, the only positively charged residue tested this way, actually resulted in an increase in the association rate.

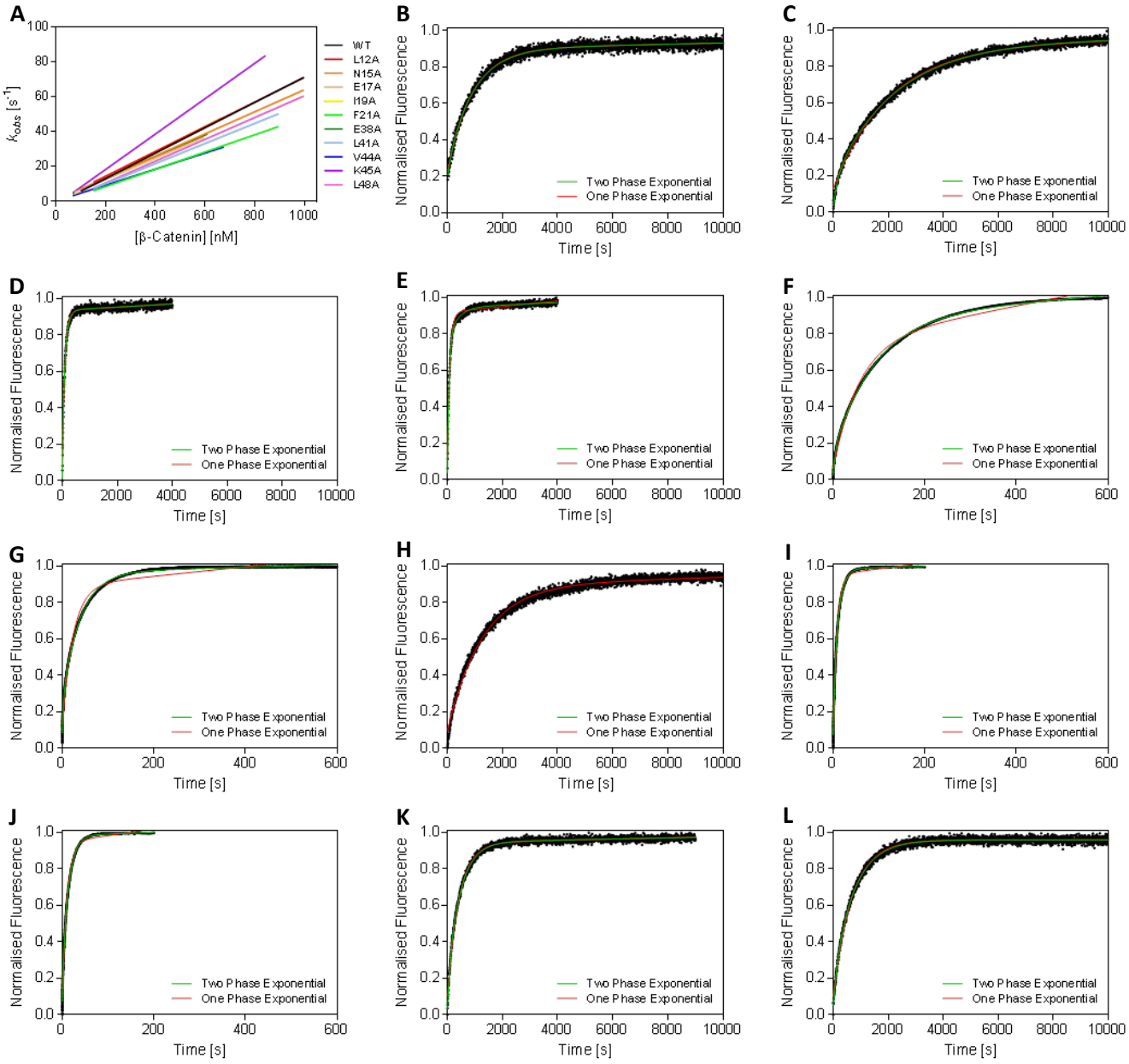


Figure 6-5 – The association and dissociation traces of the alanine scan constructs. (A) is a comparison of the association rate constants of alanine scan mutants to the pseudo wild-type construct. The individual data points have been omitted for clarity. (B) through (L) are representative dissociation traces for L12A, N15A, E17A, I19A (by fluorimeter), I19A (by stopped flow), F21A, E38A, L41A, V44A, K45A and L48A S31C constructs respectively. The dissociation traces obtain using a fluorimeter ((B) through (E), (H), (K) and (L)) and those obtained by stopped flow ((F), (G), (I) and (J)) have been on identical axes for ease of visual comparison. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

Dissociation traces were more interesting (see Figures 6-5b-l, 6-6b&c, summarised in Table 6-1), as for four of the variants (I19A, F21A, L41A and V44A) it was found that dissociation was sufficiently rapid as to be measureable on the stopped flow. Indeed for the most rapidly dissociating variants (F21A, L41A and V44A) the fluorimeter proved to be entirely unsuitable as it was only capable of taking a reading every 3 s, which would not have been sufficiently frequent to accurately capture the

faster phase. In all cases except the I19A S31C construct the dissociation was found to have a major slow phase of 70-90% and a minor fast phase of 10-30%, similar to those found for the pseudo wild-type and ensemble binding constructs (Chapter 5).

Table 6-1: Dissociation rate constants for the alanine scan constructs.

Construct	Dissociation						
	$k_{off,major}$ (x10^{-4} s^{-1})	Ratio	$k_{off,minor}$ (x10^{-3} s^{-1})	Ratio	% fast	n†	Instrument
"WT"	5.73 ± 0.40	1	1.52 ± 0.28	1	23 ± 3	8	Fluorimeter
L12A	9.77 ± 0.75	1.7	2.64 ± 0.20	1.7	19 ± 12	2	Fluorimeter
N15A	4.42 ± 0.15	0.77	3.65 ± 0.83	2.4	18 ± 2	3	Fluorimeter
E17A	89.1 ± 2.9	15.5	28.3 ± 1.9	18.7	28 ± 4	3	Fluorimeter
I19A	165 ± 7	28.8	2.92 ± 0.05	1.9	82 ± 3	2	Fluorimeter
	110 ± 0.5	19.2	207 ± 10	136	15 ± 0.1	5	SF
F21A	258 ± 2	45.0	423 ± 15	279	20 ± 0.1	8	SF
E38A	5.39 ± 0.50	0.94	2.61 ± 0.28	1.7	30 ± 1	3	Fluorimeter
L41A	993 ± 17	173	616 ± 111	406	11 ± 0.4	11	SF
V44A	742 ± 8	129	498 ± 70	328	11 ± 0.4	8	SF
K45A	21.9 ± 0.4	3.8	15.0 ± 1.3	9.9	19 ± 1	4	Fluorimeter
L48A	13.2 ± 0.1	2.3	4.24 ± 0.13	2.8	17 ± 0.6	3	Fluorimeter

† n is the number of repeats

Table 6-2: Association rate constants and $K_{d,kin}$ for the alanine scan constructs.

Construct	Association			$K_{d,kin}$ (pM)*	Ratio to WT
	k_{on} (x10^{7} M^{-1} s^{-1})	n†	Ratio to WT		
"WT"	7.33 ± 0.14	16	1	7.8 ± 0.6	1
L12A	7.19 ± 0.13	3	0.98	13.6 ± 1.1	1.7
N15A	6.25 ± 0.23	3	0.85	7.1 ± 0.4	0.91
E17A	6.30 ± 0.42	2	0.86	141.3 ± 10.5	18.1
I19A	7.04 ± 0.52	2	0.96	234.7 ± 20.0˜	30.0˜
				156.7 ± 11.6‡	20.0‡
F21A	4.82 ± 0.17	2	0.66	535.0 ± 19.1	68.4
E38A	6.20 ± 0.19	2	0.85	8.7 ± 0.8	1.1
L41A	5.95 ± 0.24	2	0.81	1668.7 ± 72.9	213.4
V44A	5.05 ± 0.46	2	0.69	1469.6 ± 135.6	187.9
K45A	10.31 ± 0.19	2	1.41	21.3 ± 0.5	2.7
L48A	5.67 ± 0.60	2	0.77	23.2 ± 2.5	3.0

* $K_{d,kin} = k_{off,major}/k_{on}$
† n is the number of repeats
˜ calculated from $k_{off,major}$ using fluorimeter
‡ calculated from $k_{off,major}$ using stopped flow

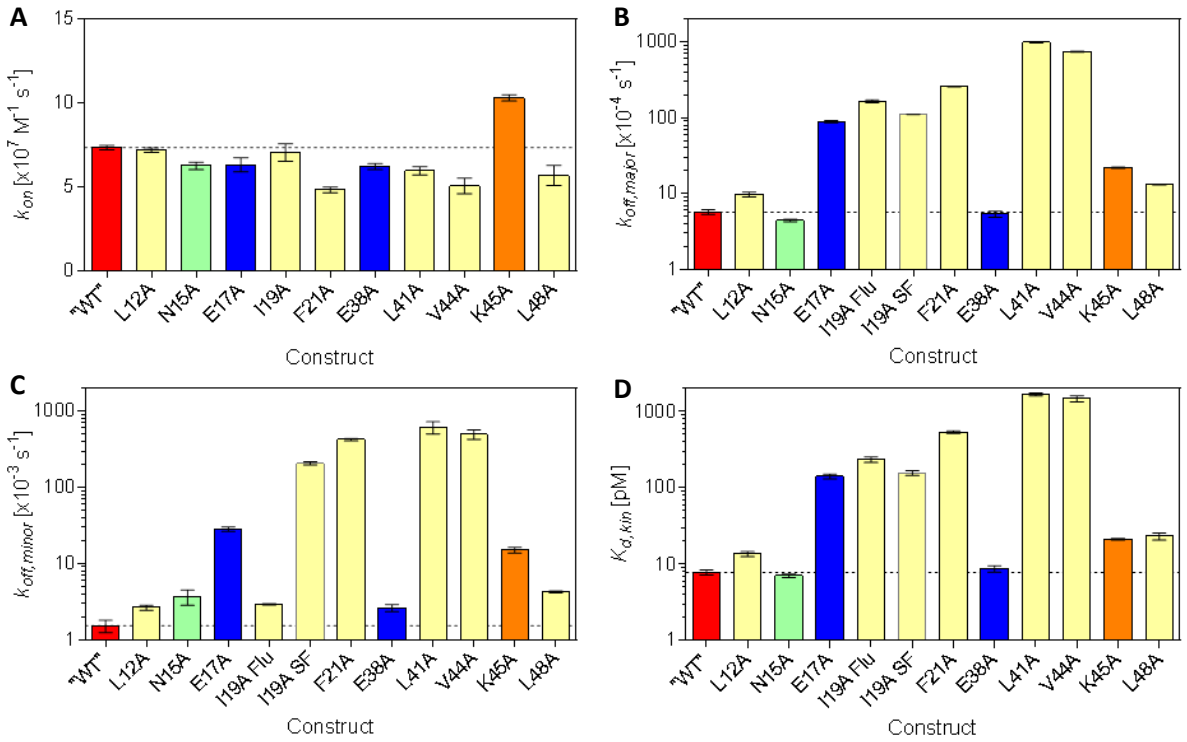


Figure 6-6 – The kinetic parameters of the alanine scan constructs. A graphical representation of the data presented in Tables 6-1 and 6-2. (A), (B), (C) and (D) are comparisons of the association rate constants, major and minor dissociation rate constants and kinetic K_d respectively for each of the alanine scan constructs. Hydrophobic residues are coloured yellow, negatively charged residue are blue, positively charge residues are orange, and polar, uncharged residues are green. All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

The I19A variant proved to be the most interesting in terms of its dissociation kinetics. The dissociation rate for this variant was such that it could be measured on both the fluorimeter and the stopped flow. When the dissociation rate constant was measured on the stopped flow the dissociation was found to be two-phase, with a major slow phase of 85% of the total amplitude and a minor fast phase of 15%, in line with the other variants. However when the dissociation rate was measured using the fluorimeter it was found to be two phase, but this time there was a major fast phase of 82% and a minor slow phase of 18%. However, the major slow phase obtained from stopped flow and the major fast phase from the fluorimeter have similar rates (1.10 x10^{-2} s^{-1} and 1.65 x10^{-2} s^{-1} respectively) and similar relative amplitudes, suggesting that association is actually three phase.

$$y(t) = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} + A_3 e^{-k_3 t} + bt + c \qquad [6\text{-}1]$$

When the individual dissociation traces for I19A S31C in both the stopped-flow and fluorimeter were fit to a three phase exponential function (Equation 6-1, see Figure 6-7), it was found that there were differences between the rate constants produced by each method (see Table 6-3).

Table 6-3: Three-phase dissociation of I19A flS31C

Instrument	Dissociation*								
	k_1	%$_1$	k_2	%$_2$	k_3	%$_3$	k_4	%$_4$	n[†]
Fluorimeter	-	-	87.6 ± 18.8	10 ± 0.5	15.0 ± 0.8	75 ± 4.5	2.66 ± 0.06	15 ± 4	2
Stopped Flow	535 ± 70[‡]	6 ± 0.5	65.2 ± 7.3	15 ± 0.8	9.1 ± 0.2	79 ± 0.3	-	-	5

* Where k_x is the xth association rate constant in x10^{-3} s^{1} and %$_x$ is the corresponding amplitude of that phase

[†] n is the number of repeats

[‡] This value is close to upper limit of what the stopped flow can measure given the dead time is ≈ 1-2 ms

For the dissociation traces produced using the fluorimeter, the major phase (75%) was the middle phase, with a rate constant of 15.0 ± 0.8 x10^{-3} s^{-1}, similar to the major phases found for a two-phase exponential fit. A minor (15%) slow phase of 2.7 ± 0.1 x10^{-3} s^{-1}, similar to the minor slow phase when from the fluorimeter, and a minor (10%) fast phase of 88 ± 19 x10^{-3} s^{-1}, which is slower than the minor fast phase from the stopped flow two-phase exponential fit, were also found.

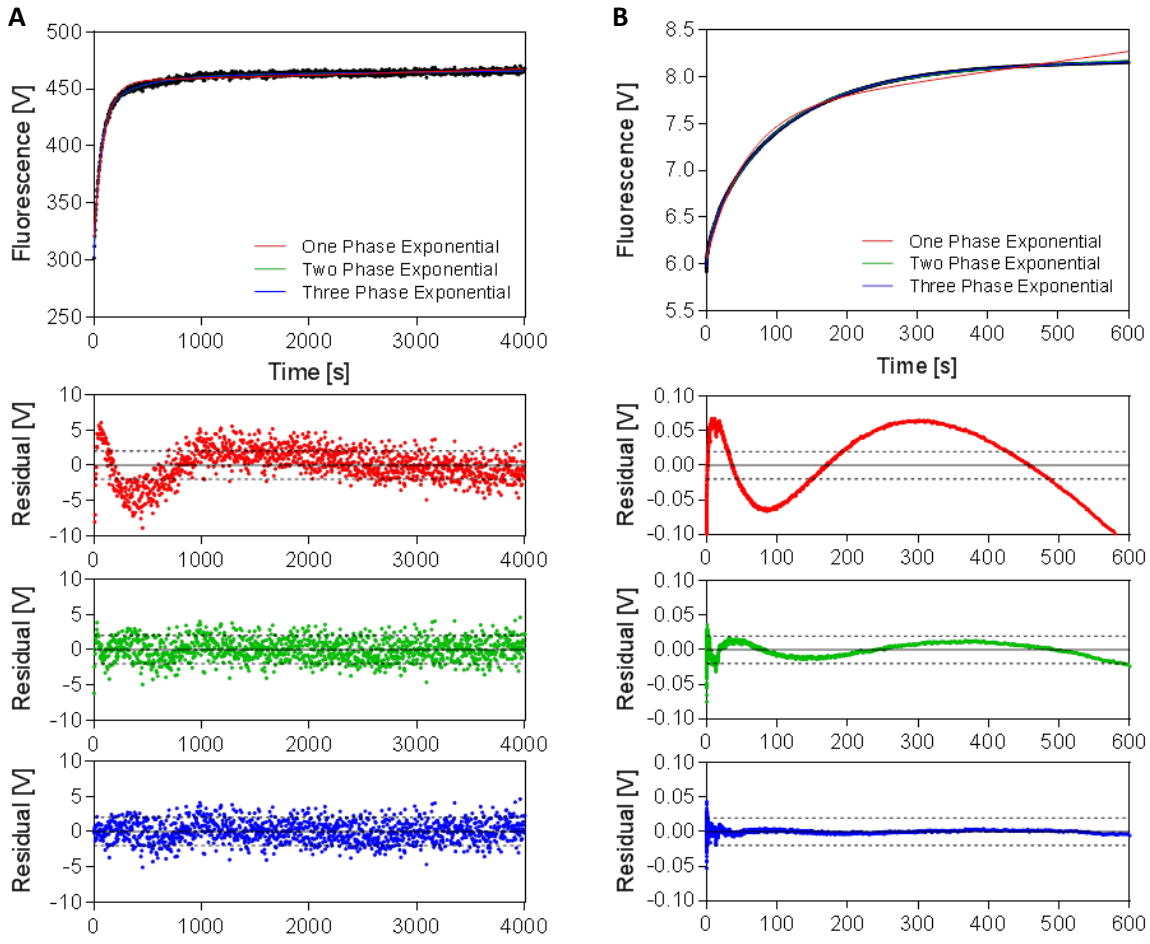


Figure 6-7 – Dissociation trace of the I19A S31C construct. (A) and (B) are representative dissociation traces for the I19A S31C construct when measured using a fluorimeter and the stopped flow respectively. These data have been plotted with one-, two- and three-phase exponential functions and their residual traces have been plotted below (red, green and blue respectively). All experiments were performed in 1x PBS, 1 mM DTT at 15°C.

For the dissociation traces produced using the stopped flow, the major phase (79%) was actually the slow phase but the rate constant was 9.1 ± 0.2 x10^{-3} s^{-1}, again similar to the previously major phases from all other fits. The minor fast(er) phase (6%) had a rate constant of 535 ± 70 x10^{-3} s^{-1}, 2.5 times that of the minor fast phase from the stopped flow, and the minor middle phase (15%) had a rate constant of 65.2 ± 7.3 x10^{-3}. This middle phase lies between the slower and faster phases of the two-phase exponential fit and is similar to the fast phase obtained for the three phase fit of the fluorimeter dissociation. Taken together these results suggest that there are a total of four phases of dissociation.

An attempt was made to fit the I19A dissociation data to a four-phase exponential function (Equation 6-2), using 0.5, 0.07, 0.01 and 0.003 s^{-1} as initial values for the rate constants and 5, 10, 75 and 10 % as initial values for the relative amplitudes. None of the seven traces could be fit unambiguously to this equation.

$$y(t) = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} + A_3 e^{-k_3 t} + A_4 e^{-k_4 t} + c \qquad [6\text{-}2]$$

The alanine scan variants can be grouped into four broad categories: those with no effect on binding (N15A and E38A and also the seven ensemble binding variants from Chapter 5), those with a mildly destabilizing effect on binding (L12A, K45A and L48A), those with a moderately destabilising effect on binding (E17A, I19A and F21A), and those with a strongly destabilising effect on binding (L41A and V44A).

In general, we can see that a significant proportion of the binding strength is provided by the interactions of hydrophobic residues (I19A, F21A, L41A and V44A). It can also be seen that there are other key salt bridges (e.g. E^{17} binding to lysine-508) in addition to the previously reported "lysine button" interactions make noticeable contributions to TCF7L2-β-catenin binding.

6.3 Φ-Value Analysis

Φ-value analysis is a protein engineering technique, developed by the Fersht Lab, for analysing the transition state of protein folding (Matouschek *et al.*, 1989; Fersht, Matouschek and Serrano, 1992; Fersht and Sato, 2004). The technique involves measuring the effect of a mutation on both the rate of folding (k_{fold}) and equilibrium stability (K_d), calculating the Gibbs free energy of the transition states and native states, relative to the denatured state, (Equations 6-3 and 6-4 respectively) for both the wild-type and variant, and calculating the ratio of the difference between the two (Equation 6-5).

$$\Delta G^{\ddagger-D} = RTln(k_{fold}) \quad [6\text{-}3]$$

$$\Delta G^{N-D} = -RTln(K_d) \quad [6\text{-}4]$$

Where R is the molar gas constant and T is the temperature in K.

$$\Phi = \frac{(\Delta G_{WT}^{\ddagger-D} - \Delta G_{mut}^{\ddagger-D})}{(\Delta G_{WT}^{N-D} - \Delta G_{mut}^{N-D})} = \frac{\Delta\Delta G^{\ddagger-D}}{\Delta\Delta G^{N-D}} \quad [6\text{-}5]$$

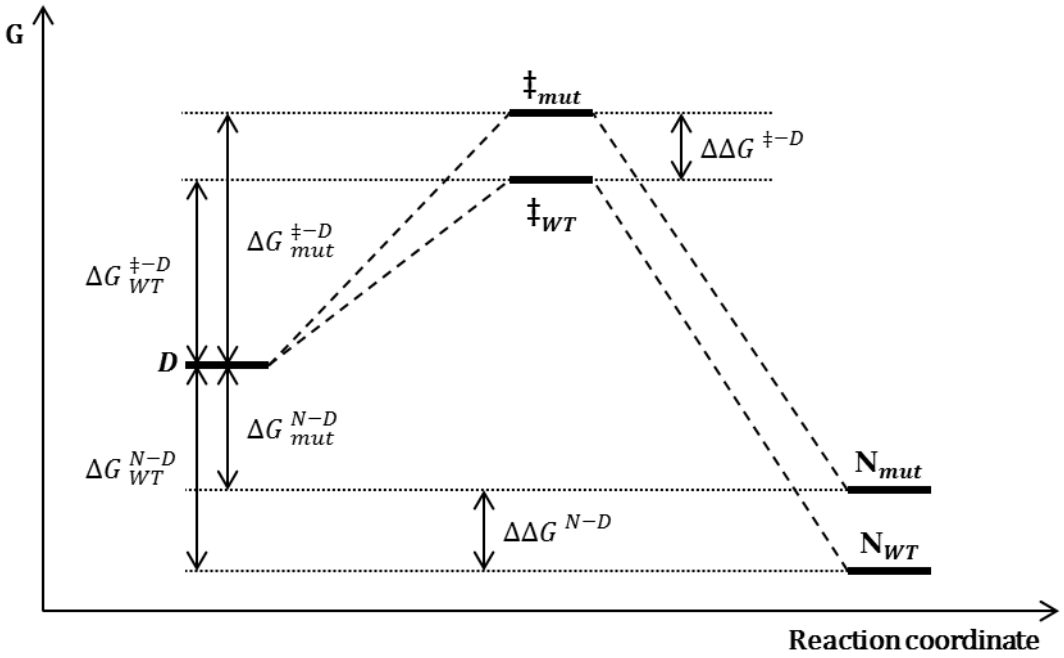


Figure 6-8 – Free energy diagram of a protein folding reaction. A free energy diagram showing the relative energy levels of the denatured (D), native (N) and rate-limiting transition state (‡) of both a wild-type protein, and the same protein with a destabilising mutation. The denatured states of both proteins have been overlaid for clarity.

Typically Φ-values fall between 0 and 1, where 0 means that the mutation has no effect on the energy of the rate-limiting transition state, i.e. that that region is unstructured in the transition state, and 1 means that the mutation has an equal effect on the energy levels of the rate-limiting transition state and the folded state, i.e. that that region is fully formed in the transition state. An intermediate value is generally interpreted as the proportion of the native contacts that that residues forms, which are formed in the rate-limiting transition state, i.e. Φ = 0.3 is interpreted as 30% of native state contacts are already formed in the rate-limiting transition state.

Non-classical Φ-values, Φ > 1 or negative Φ, have also been reported. Φ > 1, where the energy of the transition state is perturbed more than that of the native state, is usually attributed to the formation of transient contacts in the rate-limiting transition state that are not present in the native state. These are most often seen for charged residues that form few contacts in the native state and are instead solvent exposed. Negative Φ-values arise from the rare occasion where a mutation

stabilises the rate-limiting transition state but destabilises the native state, or *vice versa*. These are usually due to the elimination or introduction of a steric clash, electrostatic repulsion or hydrophilic-hydrophobic clash that occurs in one state but not the other, and are nearly impossible to interpret without further contextual information.

An analogous approach can be used for protein-protein and protein-ligand interactions, substituting the change in rate of complex formation, $\Delta\Delta G^{0\ddagger}$, and complex stability, $\Delta\Delta G^{0}$, for protein folding rate ($\Delta\Delta G^{\ddagger-D}$) and protein stability respectively ($\Delta\Delta G^{N-D}$). I applied this approach to my alanine scan variants however, it is generally required that $\Delta\Delta G^{0\ddagger}$ and $\Delta\Delta G^{0}$ be calculated separately using independent kinetic and equilibrium experiments. As I was unable to obtain meaningful results from my equilibrium binding experiments i.e. fluorescence anisotropy titration experiments (see Section 4.5), I was not able to calculate Φ-values for my alanine scan variants in the above way. Instead I have substituted my apparent $K_{d,kin}$ for the $K_{d,eq}$ to obtain estimates for the Φ-values.

Fersht and Sato (2004) suggest that Φ-values require a $\Delta\Delta G^{N-D} > 0.6$ kcal mol^{-1} (or $\Delta\Delta G^{0}$ in this case) in order to be considered non-artefactual. From this we can see that three of the residues chosen, L^{12}, N^{15} and E^{38}, cannot be analysed in this way and K^{45} should also be excluded (see Table 6-4). However, as the value for K^{45} is on the borderline, and this analysis is only qualitative, I have included it in this analysis.

Table 6-4: Φ-value analysis of the alanine scan constructs

Construct	k_{on} (x10^7 M^{-1} s^{-1})	$\Delta\Delta G^{0\ddagger}$ (kcal mol^{-1})	$K_{d,kin}$ (pM)*	$\Delta\Delta G^{0}$ (kcal mol^{-1})*	Φ*
"WT"	7.33 ± 0.14	-	7.8 ± 0.6	-	-
L12A	7.19 ± 0.13	0.01 ± 0.01	13.6 ± 1.1	0.32	n.d.
N15A	6.25 ± 0.23	0.09 ± 0.02	7.1 ± 0.4	-0.06	n.d.
E17A	6.30 ± 0.42	0.09 ± 0.04	141.3 ± 10.5	1.66	0.05
I19A	7.04 ± 0.52	0.02 ± 0.04	234.7 ± 20.0† 156.7 ± 11.6‡	1.95† 1.71‡	0.01† 0.01‡
F21A	4.82 ± 0.17	0.24 ± 0.02	535.0 ± 19.1	2.42	0.10
E38A	6.20 ± 0.19	0.10 ± 0.02	8.7 ± 0.8	0.06	n.d.
L41A	5.95 ± 0.24	0.12 ± 0.03	1668.7 ± 72.9	3.07	0.04
V44A	5.05 ± 0.46	0.21 ± 0.05	1469.6 ± 135.6	3.00	0.07
K45A	10.31 ± 0.19	-0.20 ± 0.01	21.3 ± 0.5	0.57	-0.34
L48A	5.67 ± 0.60	0.15 ± 0.06	23.2 ± 2.5	0.62	0.24

* These values are qualitative as they are calculated for $K_{d,kin}$ rather than $K_{d,eq}$ as intended.
† calculated from $K_{d,kin}$ using fluorimeter
‡ calculated from $K_{d,kin}$ using stopped flow

Overall we see that the majority of the residues tested (see Table 6-4) have yet to form bonds to β-catenin in the rate-determining transition state (Φ ≈ 0) with the C-terminal-most residue, L^{48}, having

some partially formed bonds to β-catenin at this point. K^{45} is special in that it has a negative Φ-value as a result of being faster to form the TCF7L2-β-catenin complex than the pseudo wild-type construct.

6.4 Discussion

Alanine scans are used extensively in biophysical studies, for investigating both protein folding and protein-ligand interactions. They allow us to resolve the forces that drive the formation of a specific protein's shape or complex's structure and to understand how specific dysfunctions may arise, as well as enabling us to make rational choices with predictable outcomes when manipulating them.

Understanding the kinetic pathways between folded and unfolded, or bound and unbound, has historically been a much more difficult process. Polypeptide chains sample a vast amount of conformational space in the transition from unfolded to folded or un-bound to bound state. This reaction often contains a number of energy maxima and higher energy minima which are transient and do not exist long enough to study them directly. Since its conception in the early 90's, Φ-value analysis has allowed us to describe the slowest-forming (rate-limiting) transition state and from there, we can predict the likely structures that would form between it and the native or denatured state and thereby visualise the other transition states along a folding pathway.

From the alanine scan we can see that, once again, each of the mutations has little effect on association rate (see Table 6-2), and for most of the mutations the change is less than those seen for the ensemble binding variants in Chapter 5. This is expected, as the majority of the residues tested in this chapter are hydrophobic and longer-range electrostatic interactions tend to drive protein-ligand association, as opposed to shorter-range hydrophobic interactions seen in protein folding (hydrophobic collapse). That being said, three residues stand out from the others as being worthy of note: F^{21}, V^{44} and K^{45}.

K^{45} is interesting, as it is the only variant tested that resulted in an increase in the association rate, if only a small one. It is also the only positively charged residue tested in this manner so it is easy to speculate that the cause of the increased association rate is the removal of an electrostatic repulsion. Indeed, when we examine the coulombic surface of β-catenin we can see that the groove into the C-terminal α-helix binds is flanked by multiple positively charged residues (see Figure 6-9). Further inspection of the three crystal structures of the TCF7L2-β-catenin complex shows that, when bound, TCF7L2 K^{45} does not form a more common salt bridge to a negatively charged residue on β-catenin, but forms two, more unusual, cation-π interactions with two buried histidine residues: histidine-219 and histidine-260. Due to their buried nature, these cation-π interactions are unlikely

to provide a significant long-range attractive force, so by mutating K^{45} to alanine no attractive interaction is lost, but the repulsive interactions are reduced. Hence the association rate constant increases.

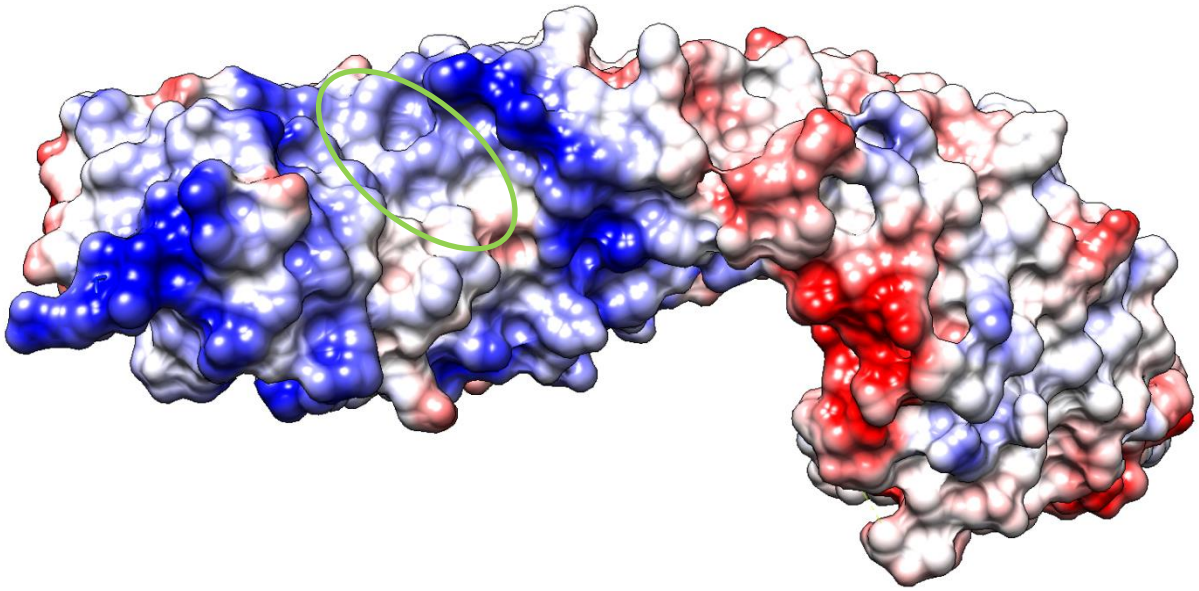


Figure 6-9 – The coulombic surface of β-catenin. The surface of β-catenin coloured by coulombic charge, with regions of positive charge coloured blue and regions of negative charge coloured red. The area circled in green indicates the groove into which the C-terminal helix of TCF7L2 binds. TCF7L2 residues V^{44} and K^{45} dock into the small pocket within the highlighted area. Structure image was created using UCSF Chimera from PDB ID 2Z6H

It is a bit more difficult to explain the effect of the V^{44} and F^{21} variants on the association rate constant. V^{44} is a bulky β-branched hydrophobic residue so loss of long-range electrostatic interactions on mutation to alanine is not a factor. It could be that steric hindrance from V^{44} limits the conformational freedom of unbound TCF7L2 increasing the likeliness of it being in a more favourable β-catenin binding confirmation. This could indicate that the α-helix is pre-formed before the rate-limiting transition state, but we do not have enough evidence to support that conclusion. Of note is the fact that V^{44} is immediately adjacent to K^{45}, whose presence has a retarding effect on the association rate. It could also be that the bulkiness of V^{44} forces K^{45} to adopt less disfavourable conformations and by mutating to it to alanine, it removes this interaction.

For the F^{21} variant it is less clear why there is a comparatively large effect on association rate but it is likely due to this residue's interaction with arginine-386 on β-catenin. In all three crystal structures, F^{21} forms an interaction with this arginine that can be considered both a π-stacking interaction and a cation-π interaction and is likely a major factor in how much F^{21} contributes to TCF7L2-β-catenin binding (see Figure 6-10). While cation-π interactions are not as effective at longer ranges as electrostatic interactions are, it is likely that arginine-386 does provide an attractive force on F^{21},

hence mutating F^{21} to alanine has an effect on association, unlike the buried histidines which interact with K^{45}.

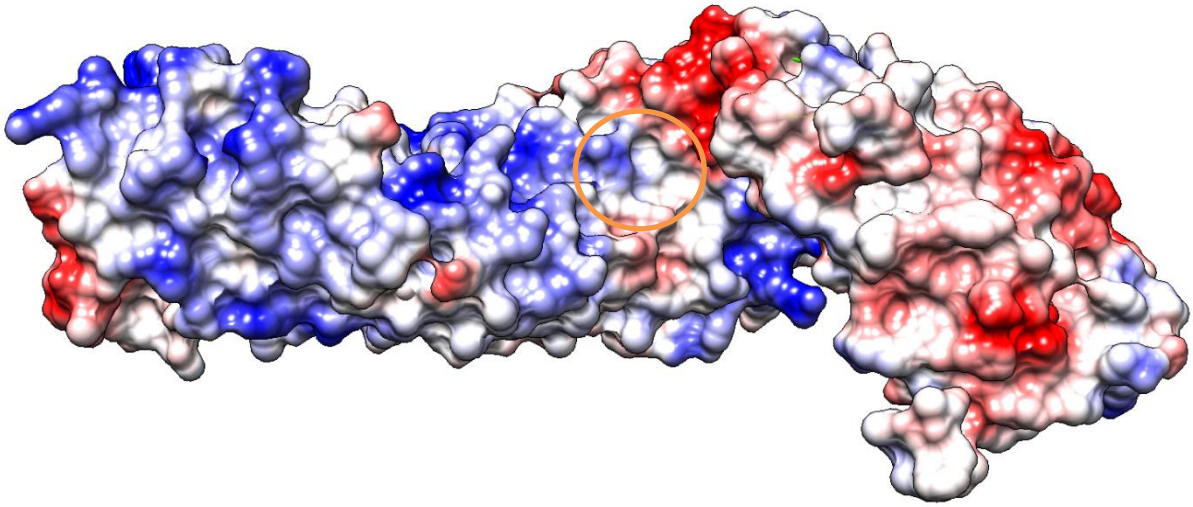


Figure 6-10 – Rotation of Figure 6-9 to show phenylalanine-21 binding pocket. The same surface of β-catenin as seen in Figure 6-9, but rotated to show the F^{21} binding region. The area circled in orange indicates pocket into which F^{21} binds along with β-catenin arginine-386 to which F^{21} bonds. Structure image was created using UCSF Chimera from PDB ID 2Z6H.

Unlike the association rate constants, the dissociation rate constants were significantly more perturbed in the alanine scan, with changes observed by as much as two orders of magnitude from wild-type (see Table 6-1). We can see that those residues which vary the most significantly between the three crystal structures i.e. L^{12}, N^{15} and E^{38}, perturb dissociation rate the least. This makes sense because if these residues are seen varying between structures, it implies their interactions are weaker and more transient, hence the small increase in dissociation rate. Conversely, the hydrophobic residues i.e. I^{19}, F^{21}, L^{41} and V^{44}, resulted in the largest increases in the dissociation rate constants, implying that these residues form very strong interactions with β-catenin residues. This is reflected by these residues being invariant between crystal structures.

Mutating E^{17} to alanine also has a significant effect on dissociation, unlike the other negatively charged residues tested (E^{38} and those from Chapter 5). This may be because there is a specific positively charge pocket into which E^{17} can dock, and form specific salt bridges to β-catenin residues arginine-469, histidine-470 and lysine-508. We can also see that the neighbouring residue, D^{16}, sits in an adjacent, deeper, more positively charge pocket (see Figure 6-11). I had opted not to test this residue as other papers have shown that the equivalent of this interaction has a very significant effect on β-catenin's other binding partners, and from this we can see why.

The I19A mutation invokes a lot of discussion. The major phase dissociation rate ended up being one order of magnitude different from the wild-type with a half-life of approximately 60 s. This was fast enough to allow the recommended ten half-lives to be recorded on the stopped flow (the upper

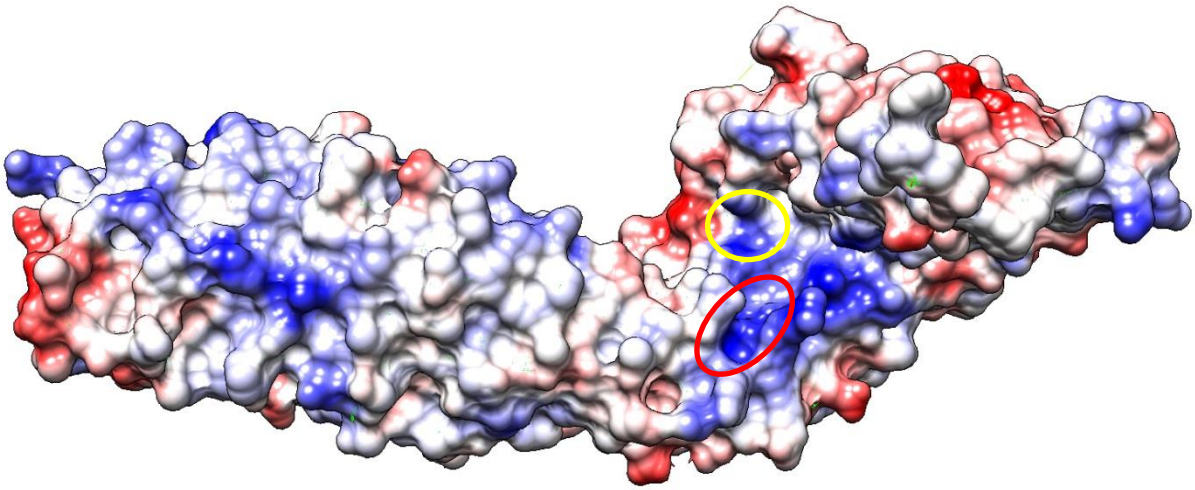


Figure 6-11 – Rotation of Figure 6-9 to show glutamate-17 and aspartate-16 binding pockets. The same surface of β-catenin as seen in Figure 6-9, but rotated to show the E^{17} & D^{16} binding region. The area circled in yellow indicates the positively charge pocket into which E^{17} binds and the red indicates the positively charge pocket into which D^{17} binds. Structure image was created using UCSF Chimera from PDB IDs 2Z6H.

time limit for that machine being 600 s), while at the same time slow enough to be accurately captured on the fluorimeter, which can take a reading every 3 s (20 points per half-life).

The fact that both techniques produce near identical $k_{off,major}$ values but different $k_{off,minor}$ values, one an order of magnitude faster and the other an order of magnitude slower than $k_{off,major}$, strongly indicates the existence of a third phase of dissociation (see Table 6-1). Then when these traces were fitted to a three phase decay it was found that the two techniques produced two phases that were similar between the two techniques and each had a unique third phase, giving us four distinct phases (see Table 6-3). Neither data set could be fit to a four phase decay. This is likely due to the difference in time scale for the techniques and the difference in magnitude of the phases. The four phases of approximately 0.5, 0.075, 0.01 and 0.003 s^{-1} have corresponding half-lives of 1.4, 9.2, 69 and 230 s respectively. The fastest phase is too fast to accurately captured on the fluorimeter, which takes a reading every 3 s, and the second fastest phase would not be described particularly accurately, as is demonstrated in the comparatively high error in the value. Conversely the slowest phase is too slow to be able to capture ten half-lives on the stopped-flow, with a maximum recording time of 600 s, and is not quite able to capture ten half-lives of the second slowest phase.

These additional dissociation phases are not unique to the I19A S31C variant; a similar observation can be made for the dissociation trace for the F21A S31C variant too. When the stopped flow dissociation data are fit with a two-phase exponential this gives a major slow phase (80%) of 25.8 ± 0.2 $x10^{-3}$ s^{-1} and a minor fast phase (20%) of 423 ± 15 $x10^{-3}$ s^{-1}. But when these data were fit to a three-phase decay a major slow phase (66%) of 20.3 ± 0.2 $x10^{-3}$ s^{-1}, a minor middle phase (22%) of 91.8 ± 5.9 $x10^{-3}$ s^{-1}, and a minor fast phase (12%) of 794 ± 68 $x10^{-3}$ s^{-1} were produced. Again these data could not be fit to a four-phase decay. This trend in rate constants is similar to those of the

stopped-flow dissociation traces of the I19A S31C construct. In contrast, the stopped flow dissociation traces for the L41A S31C and V44A S31C constructs fit best to a two-phase exponential and could not be fit with a three-phase exponential.

It is curious that three- or four-phase decays can be observed only for the I19A S31C and F21A S31C constructs and not for the other fast-dissociating variant constructs. The fact that the mutations that show this behaviour are located in the N-terminal binding region and those that do not are located in the C-terminal binding region, suggests this phenomenon might be a unique property of the N-terminal binding region. If so, then a potential reason for this behaviour could be the neighbouring variable region, and what is being observed might be multiple conformational rearrangements of the ensemble binding region, as several of the more highly populated conformations are also perturbed by mutations within the N-terminal binding region.

The half-life of the major phase of the F21A S31C construct is 27 s which I believe would be slow enough to be accurately captured on the fluorimeter (nine points per half-life). If there is a fourth phase then I would expect to see a two-phase decay, with a major fast phase with a rate constant similar to the major phase obtained from fitting the stopped flow dissociation traces, and a minor slow phase. Similarly the half-life of the major phase of the E17A S31C construct is 78 s, which may be slow enough that it can be fitted accurately from a stopped flow trace. If this four-phase theory is true I would expect to see a three-phase decay with a major slow phase and minor middle and fast phases. Unfortunately I was unable to perform these experiments due to time constrains.

The trend in $K_{d,kin}$ is that the C-terminal region of TCF7L2 appears to be contributing more energy towards the TCF7L2-β-catenin interaction than the N-terminal region. However, due to the invariance in association rates constant between alanine variants, the trends in $K_{d,kin}$ are a reflection of the trends seen in the major dissociation rate constants. As a result of this, when estimating the Φ-values as I have done, we see that most of the residues tested that fulfil the recommended $\Delta\Delta G^{N-D}$ ($\Delta\Delta G^{0}$) > 0.6 kcal mol^{-1} have Φ-values close to zero, i.e. essentially unformed in the rate-determining transition state. The exceptions are K^{45} which has a negative Φ-value and L^{48} for which Φ ≈ 0.25.

The negative Φ-value for K^{45} is much more difficult to interpret, as the magnitude of ≈ 0.35 could be an artefact as a consequence of the negative $\Delta\Delta G^{\ddagger-D}$. If the $\Delta\Delta G^{\ddagger-D}$ of -0.2 kcal mol^{-1} was the sum of a 0.2 kcal mol^{-1} loss of favourable interactions plus a 0.4 kcal mol^{-1} loss of disfavourable interactions, then the magnitude of all changes on mutation to alanine, is 0.6 kcal mol^{-1}. This would then give Φ ≈ 1.

$\Phi \approx 0.25$ compared to ≈ 0 for most of the other residues, suggests that L^{48} makes contact with β-catenin at an earlier stage in the association pathway that the other residues. This is suggestive of a "zipping up" binding mechanism, where the C-terminal region of the long, string-like TCF7L2 binds to β-catenin first and the remaining residues follow in succession as the enforced proximity promotes binding. Also of note is that the estimated Φ-values are not all zero across the length of the C-terminal α-helix, suggesting the α-helix is not yet formed in the rate-determining transition state.

Chapter 7 - Conclusions

Many of the diseases that are the most resistant to treatment are challenging because of the complexity of their underlying biochemical and pathological processes. Diseases like cancer and Parkinson's are difficult to treat as there are multiple pathways that can lead to the disease. In this Thesis I aimed to provide new information about the canonical Wnt signalling pathway, which underlies both these diseases.

The gene PARK8, which encodes for the protein Leucine-rich repeat kinase 2 (LRRK2), is one of just a small number of genes with a demonstrable link to Parkinson's disease (Drolet, Sanders and Kern, 2011). Even though this gene accounts of only 5% of familial and 3% of sporadic Parkinson's disease cases (Tan and Jankovic, 2006), it is still the highest single attributable cause of Parkinson's disease, and therefore represents a major target for the treatment of the disease.

Despite years of intensive study, LRRK2 has proved to be an elusive target. It has been shown to interact with a wide variety of intracellular partners including those involved in cytoskeleton dynamics, mitochondrial homeostasis, the MAPK/ERK pathway, and the canonical Wnt signalling pathway (see Section 1.4, Kang and Marto (2017)). Yet, without any concrete structural data for the LRRK2 protein further, information about the nature of these interactions will be difficult to obtain.

In this Thesis I attempted structural studies of the LRRK2 protein domains, based on the predictions of Mills *et al*. (2012). Mills *et al.* predicted through homology modelling, the existence of four tandem repeat domains in the unsolved regions of LRRK2. This may represent a significant finding as tandem repeat domains typically facilitate protein-protein interactions by providing a stable and specific architecture for other protein partners to interact with (Grove, Cortajarena and Regan, 2008). Unfortunately, after a year of using a variety of different approaches I was unable to solubly express these domains using recombinant expression in *E. coli*. This in no way disproves the predictions of Mills *et al.*, only that the approach I took in trying to isolate the individual domains was not appropriate. Other research groups around the world have been able to express full-length human LRRK2 primarily in human cell lines; however expression in human cell lines does not tend to produce proteins in the quantities required for the biophysical assays I had intended and so this project was brought to a conclusion.

The other focus of this Thesis is the interaction of β-catenin with the transcription factor TCF7L2. As the signal transducing species in the canonical Wnt signalling pathway, β-catenin plays an important role in both embryonic development and the maintenance of adult tissues, and consequently its

dysfunction has been implicated in a wide range of cancers (MacDonald, Tamai and He, 2009). It is therefore an attractive target in "the fight against cancer". However as part of its role in the canonical Wnt signalling pathway, as well as its role at adherens junction, β-catenin interacts with a larger number of protein partners, many of which share the same binding epitope, i.e. the inner groove of the ARM repeat superhelix. This feature makes it a challenging target for the development of therapeutic agents, as disruption of one interaction may have undesirable side-effects as a result of disrupting other interactions. The research I have conducted and presented here is part of a larger ongoing project in the Itzhaki group to understand the mechanism of the interactions of β-catenin with its multiple binding partners, with Dr S.D. Dunbar focusing on β-catenin's interactions with Axin and E-cadherin, and Dr P.J.E. Rowling focusing on the APC-β-catenin interactions.

When I began this project, there existed three unique crystal structures of TCF7L2 bound to β-catenin (Graham *et al.*, 2001; Poy *et al.*, 2001; Sampietro *et al.*, 2006), which differed in terms of which glutamate residue (E^{24} or E^{29}) of TCF7L2 formed a salt bridge with the lysine-312 of β-catenin. I showed that mutating these glutamate residues to alanine, as well as other negatively charged residues in that region, had a negligible effect on the association and dissociation kinetics, at a physiologically relevant osmolarity (Section 5.4). However, in later experiments (Section 5.5) significant differences were seen in the dissociation kinetics between the pseudo wild-type construct ("WT" S31C) and the construct with five negatively charged residues mutated to alanine (XXXXX S31C) at higher osmolarities.

These results imply that this region of TCF7L2 (residues K^{22} to D^{39}), which I refer to as the "variable" region, does interact with β-catenin but, at the physiologically relevant osmolarity under which I first studied it, these interactions make little contribution to the overall binding affinity. The simplest conclusion that can be drawn from these data is that the differences that are visible in the three crystal structures are consequence of the different conditions in which the proteins were crystallised, including varying osmolarity, which seems to have a significant effect on this region and is therefore likely reflected in the observed structures (Das and Pappu, 2013). However, as I already noted in Section 5.5, this does not mean that these are not real structures that TCF7L2 is capable of adopting when in complex with β-catenin in solution, but that these structures are in constant flux and these observed crystal structures are snapshots of a dynamic system.

Charge appears to play a significant role in β-catenin's binding to its other partners as well. For both APC and E-cadherin, phosphorylation significantly increased their binding affinities for β-catenin (310-fold and 880-fold respectively Choi, Huber and Weis (2006)). Conversely, LEF1 which has higher

homology with TCF7L2 then APC or E-cadherin, can also be phosphorylated *in vivo* but does not gain increased binding affinity for β-catenin (Sun and Weis, 2011).

When we examine the crystal structures of each of these partners (1TH1, Xing *et al.*, (2004), 3IFQ, Choi *et al.*, (2009), and 3OUJ, Sun and Weis, (2011) respectively) we get a mixed picture. E-cadherin, which has the greatest increase in binding affinity, is phosphorylated at three residues, in equivalent positions to residues E^{26}, E^{28} and the other within the C-terminal helix of TCF7L2**.** APC, which also gains a significant increase in affinity, is phosphorylated at five residues, one in an equivalent position to residue E^{17} and four within the C-terminal helix of TCF7L2. However LEF1 is phosphorylated at just one residue, in an equivalent position to residue A^{37} which doesn't increase the binding affinity significantly. These data suggest that it is phosphorylation within the equivalent of the C-terminal helix that results in the increased affinity, not the variable region. This may account for the change in kinetics I observed when I tried labelling TCF7L2 at position S^{46}.

I next investigated the contribution of the other TCF7L2 residues to the TCF7L2-β-catenin binding interaction and kinetics by means of an alanine scan (Chapter 6). The collected equilibrium data from the four sets of alanine scans, three previously published (Omer *et al.*, 1999; Fasolini *et al.*, 2003; Gail, Frank and Wittinghofer, 2005) and my own, are shown in Table 7-1. For ease of comparison between data sets the ratio between the values obtained for the alanine scan variants with their respective wild-type constructs has been calculated.

From both my own alanine scans and those from the literature we can see a trend in the residues (see Table 7-1). There are two regions of TCF7L2, which I will call the N-terminal binding region (residues N^{16} to F^{21}) and the C-terminal binding region (residues L^{41} to L^{48}) that contribute significantly more to the binding affinity than other regions and these are the regions that are the most similar between crystal structures (see Figure 7-1). Targeting these regions in particular may prove to be a good starting point for the rational design of novel therapeutic peptides or

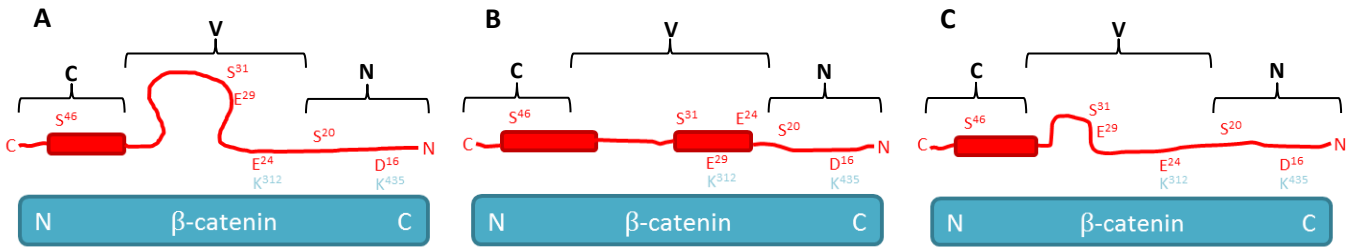


Figure 7-1 – The three regions of the TCF7L2. Cartoon diagrams of the three crystal structures of TCF7L2 in complex with β-catenin with the three regions of TCF7L2: the variable domain (V), and highlighted. (A), (B) and (C) are based on the crystal structures obtained by Poy *et al.*, Graham *et al.* and Sampietro *et al.* respectively (see Chapter 4). For all three structures the N-terminal binding domain (N) and the C-terminal binding domain (C) are very similar, but the variable region (V) adopts alternative structures in each crystal structure.

peptidomimetic drugs that target the TCF7L2-β-catenin interaction.

Table 7-1: Comparison of literature alanine scans and my own alanine scan data.

Data Set	My Data		Omer *et al.*[1]		Fasolini *et al.*[2]		Gail *et al.*[3]	
Construct	1-54 S31C		1-100		1-56		1-53	
Method	Kinetics		ELISA		ITC		Fluorescence Polarisation	
Variant	$K_{d,kin}$ (pM)*	**Ratio**	$K_{d,eq}$ (pM)	**Ratio**	$K_{d,eq}$ (pM)	**Ratio**	$K_{d,eq}$ (pM)	**Ratio**
Wild-type	7.8 ± 0.6	1.0	23 ± 3 x10^3	1.0	3.6 ± 1.9 x10^3	1.0	840.0 ± 180	1.0
D10A	-	-	-	-	2.2 ± 1.2 x10^3	0.6	-	-
D11A	-	-	-	-	20 ± 40 x10^3	5.6	-	-
L12A	13.6 ± 1.1	1.7	-	-	-	-	820.0 ± 130	1.0
G13A	-	-	-	-	-	-	520.0 ± 85	0.6
N15A	7.1 ± 0.4	0.9	-	-	-	-	-	-
D16A	-	-	Complete Loss	n/a	143 ± 41 x10^3	39.7	1400.0 ± 97	1.7
E17A	141.3 ± 10.5	18.1	390 ± 80 x10^3	17.0	-	-	21 ± 5.6 x10^3	25.0
L18A	-	-	-	-	3.7 ± 2.1 x10^3	1.0	12 ± 2.6 x10^3	14.3
I19A	234.7 ± 20.0	30.0	-	-	3.7 ± 1.1 x10^3	1.0	31 ± 3.2 x10^3	36.9
F21A	535.0 ± 19.1	68.4	-	-	1.9 ± 0.5 x10^3	0.5	220 ± 24 x10^3	261
D23A	5.7 ± 0.4	0.7	-	-	2.2 ± 0.6 x10^3	0.6	1.8 ± 0.6 x10^3	2.1
E24A	8.3 ± 0.7	1.1	-	-	12.5 ± 1.6 x10^3	3.5	-	-
E26A	8.5 ± 0.6	1.1	-	-	-	-	-	-
E28A	8.5 ± 0.8	1.1	-	-	2.6 ± 0.6 x10^3	0.7	-	-
E29A	15.6 ± 1.4	2.0	-	-	2.9 ± 1.6 x10^3	0.8	-	-
N34A	-	-	-	-	3.4 ± 1.0 x10^3	0.9	660.0 ± 310	0.8
E38A	8.7 ± 0.8	1.1	-	-	-	-	-	-
D40A	-	-	-	-	9.1 ± 6.6 x10^3	2.5	-	-
L41A	1668.7 ± 72.9	213	-	-	39 ± 44 x10^3	10.7	250 ± 12 x10^3	297
V44A	1469.6 ± 135.6	187	-	-	48 ± 23 x10^3	13.2	-	-
K45A	21.3 ± 0.5	2.7	-	-	4.5 ± 1.4 x10^3	1.3	-	-
S46A	-	-	12 ± 2 x10^3	0.5	-	-	-	-
S47A	-	-	20 ± 2 x10^3	0.9	-	-	-	-
L48A	23.2 ± 2.5	3.0	4 ± 0.5 x10^6	173	-	-	220 ± 22 x10^3	261
V49A	-	-	-	-	-	-	7.0 ± 7.3 x10^3	8.3
E51A	-	-	-	-	-	-	2.5 ± 0.4 x10^3	3.0
E24A E29A	6.9 ± 0.6	0.9	-	-	-	-	-	-
XXXXX	6.9 ± 0.6	0.9	-	-	-	-	-	-

$* \ K_{d,kin} = k_{off,major}/k_{on}$

[1] (Omer *et al.*, 1999); [2] (Fasolini *et al.*, 2003); [3] (Gail, Frank and Wittinghofer, 2005)

The trends in the alanine scans suggest that TCF7L2 binds to β-catenin through an avidity binding mechanism (Olsen, Teilum and Kragelund, 2017) with the two distinct binding regions (the N- and C-terminal binding regions), linked by a disordered linker, binding to two corresponding sites on β-catenin. This model would explain the apparent complexity of the binding kinetics.

For alanine variants in the C-terminal binding region, the dissociation appeared to be biphasic, which could be explained as the sequential dissociation of the two binding regions (see Figure 7-2a), with binding region one (BD1) always observed dissociating before binding region 2 (BD2). However the alanine variants in the N-terminal region appeared to produce a four-phase dissociation. A possible explanation for this is an additional alternative dissociation pathway in which BD2 now dissociates before BD1, giving two additional observable phases (see Figure 7-2b).

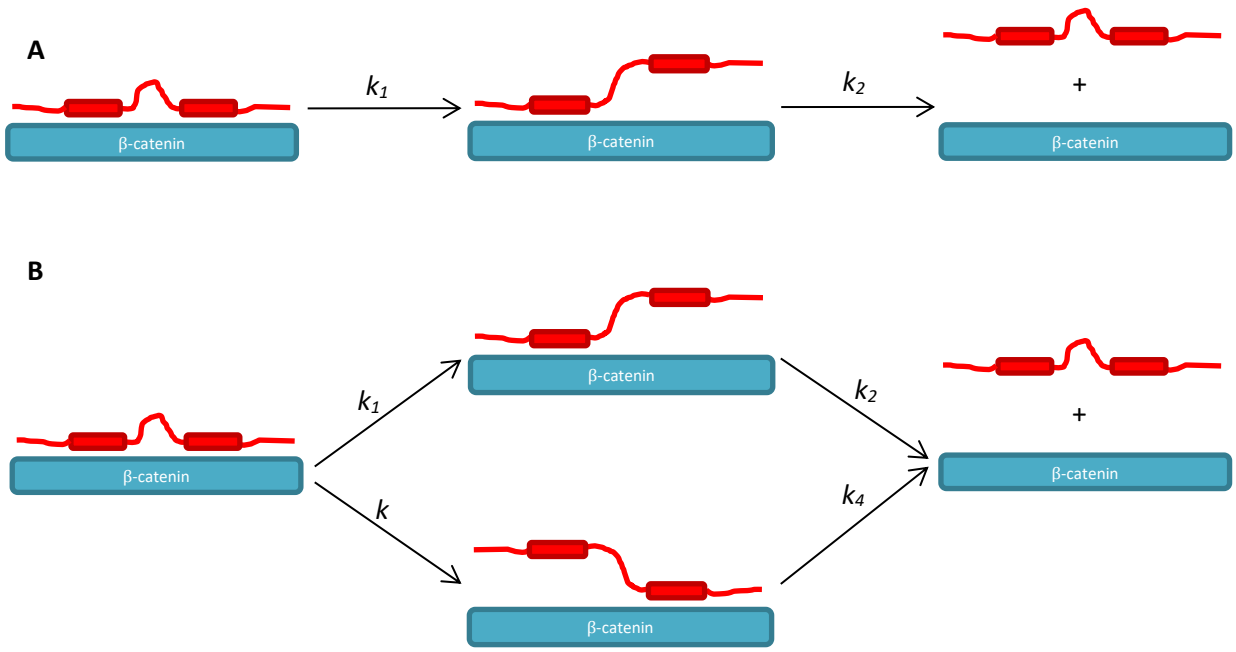


Figure 7-2 – The proposed mechanism of TCF7L2-β-catenin complex dissociation. (A) the sequential dissociation pathway for the two TCF7L2 (red) binding regions from β-catenin (cyan), in which one region always dissociates before the other region. (B) the branched dissociation for the two TCF7L2 binding regions from β-catenin, in which either of the two can dissociate first.

It is likely that this second pathway was not sufficiently populated to be detected for the wild-type construct, due to the energetics involved, and that perturbation of the binding interaction by mutating to alanine allowed flux through this second pathway. This therefore implies that it is the C-terminal binding region (41-48) which dissociates first, followed by N-terminal binding region (16-21). Mutations in the N-terminal then weaken the binding of that region, accelerating its dissociation and allowing flux through the second pathway. This is consistent with alanine variants in the C-terminal region having a greater effect on the dissociation rate constant in general than the N-terminal alanine variants.

When we look at the crystal structures of β-catenin binding to its binding partners we can see that each partner binds to a subset of four distinct regions on β-catenin (see Figure 7-3). The first, the N-Cap region, comprises the first ARM repeat as well as the capping N-helix which is present in full length β-catenin (Xing *et al.*, 2008) but is not present in the Weis construct used in this study. Second is the "phosphorylation" region, which includes ARM repeats two to four and is where the C-terminal binding region of TCF7L2 binds. This is where the phosphorylated residues of phospho-E-cadherin and phospho-APC bind. Next is the lysine button region, composed of ARM repeats five to ten which includes the lysine buttons, lysine-312 and lysine-435. Finally there is the C-cap region comprising the eleventh and twelfth ARM repeats and the capping C-helix which is also not present in the Weis construct.

Whether the two-site avidity model for binding can be applied to all of β-catenin's binding partners has yet to be determined. I would speculate, based on their sequence and structural homology, that the other members of the TCF/LEF family of the transcription factors (TCF7, LEF and TCF7L1) might exhibit a similar binding mechanism. The other binding partners are harder to predict. It is likely that two-site avidity mechanisms require a disordered linker of an appropriate length, such as the variable region of TCF7L2. In Dr. Rowling's study of the APC repeat 3 binding to β-catenin she found that association and dissociation were single phase. However when she inserted an 18-residue disordered linker in her APC construct, in an equivalent position to the TCF7L2 variable region, she found that association became two phase. These studies are currently ongoing and unpublished.

Another curiosity is that although there seems to be this multi-phasic dissociation, association seemed to be exclusively single phase. While the association traces could also be fit with a two phase exponential decay, there was no consistency in the second association rate constants phases (see Section 4.5). However when testing the "WT" S20C construct there appeared to be a different association phase that was difficult to distinguish due to a poor signal-to-noise ratio. This could be a second association phase that is just not observable for the "WT" S31C and "WT" S46C constructs due to the relative intensities of the signals. Similar to the two phase dissociation, this second association phase could be due to the sequential association of the two binding regions.

In summary it appears that, due to its intrinsically disordered nature, the TCF7L2-β-catenin kinetics are very complex. My findings demonstrate a fast association combined with a very slow, multi-phasic dissociation resulting in a more tightly bound complex than has previously been suggested. This very slow dissociation is likely to confound some equilibrium measurements of complex stability, particularly those that rely on titration (e.g. ITC), as the reaction may not have the necessary time to reach equilibrium before the system is perturbed again. Indeed this effect can be

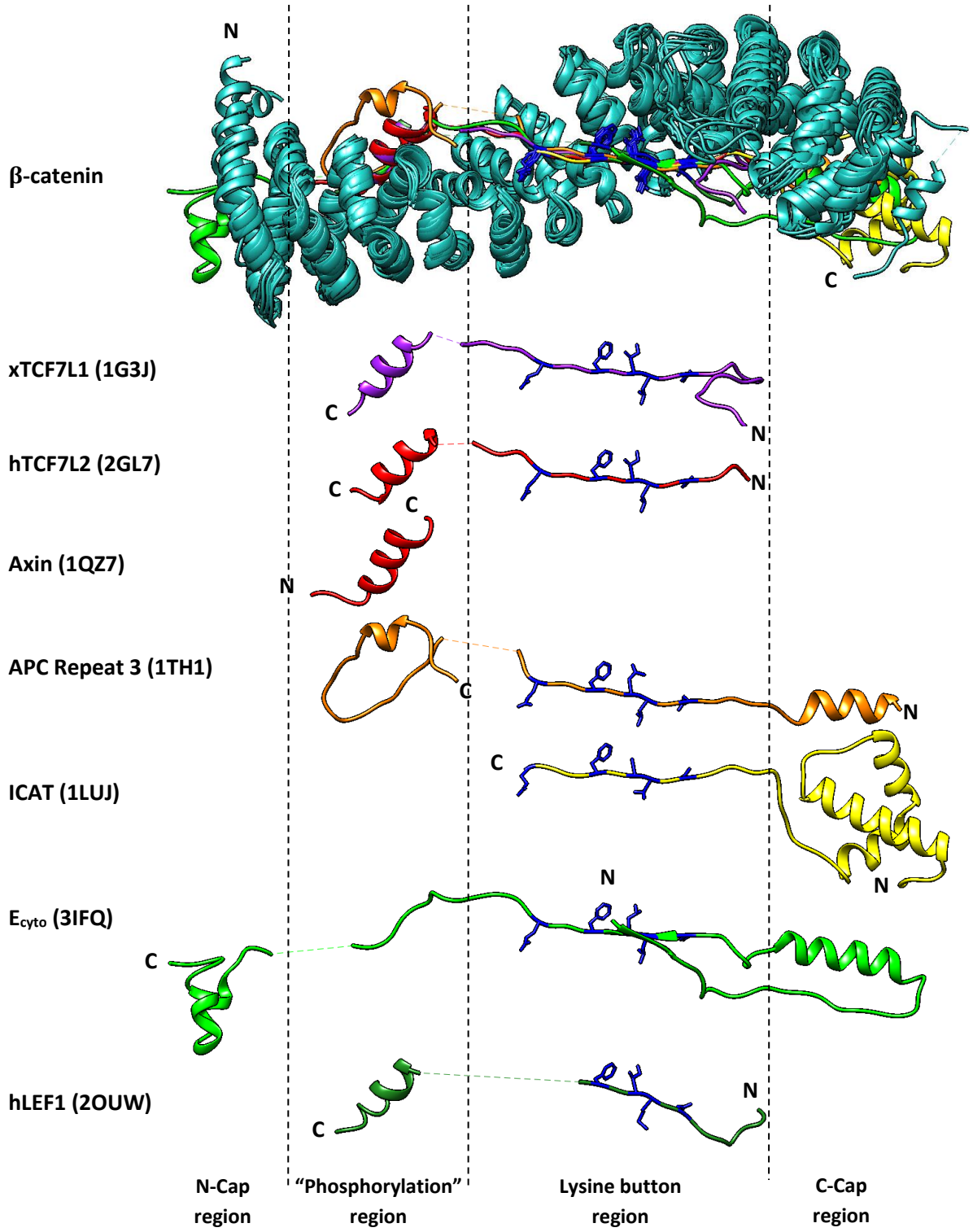


Figure 7-3 – The binding regions of β-catenin: The superposition of the crystal structures of β-catenin in complex with its binding partners as previously shown in Section 1.5.2. The four proposed binding regions have been highlighted, demonstrating the different modes in which β-catenin's partners bind.

seen from the differences in wild-type complex stability between studies (see Table 8-1, Omer *et al.*, 1999; Fasolini *et al.*, 2003; Gail, Frank and Wittinghofer 2005). My findings also suggest that the TCF7L2-β-catenin binding interaction follows a two-site avidity binding mechanism which future studies will need to account for.

7.1 Unfinished Work

There are also a number of experiments which I had begun working on which I was unable to complete due to time constraints. First I began investigating a way of actually forcing TCF7L2 formation of the variable α-helix seen only in Graham *et al.*'s structure, as well as attempting to determine the extent to which the C-terminal α-helix is pre-formed, by means of peptide stapling. Peptide stapling is the process of macrocyclising a peptide via unnatural amino acid side chains, typically across one face of an α-helix (Walensky and Bird, 2014). This technique has even be previously used in targeting β-catenin although never using TCF7L2 (Grossmann *et al.*, 2012; Takada *et al.*, 2012).

7.1.1 Peptide Stapling

Previous studies targeting β-catenin used a simple hydrocarbon staple formed from a ruthenium-catalysed (Grubbs' catalyst) ring-closing metathesis (RCM) reaction of two alkene containing side chains (see Figure 7-4a). However the hydrocarbon staple is quite hydrophobic which can result in a change in binding conformation to compensate for this new interaction, as well as reduced solubility. Previous experiments with stapled peptides in our lab have used click chemistry (Kolb, Finn and Sharpless, 2001) as a method for forming a peptide staple (Guttenplan *et al.*, 2017; Xu *et al.*, 2017). This process uses the Copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction described by Himo *et al.* (2005) to form a double-clicked staple between two azide containing unnatural amino acid residues and a dialkyne linker (see Figure 7-4b). This has the advantage over RCM of forming two, more soluble triazole rings instead of an alkene as well as allowing additional functionalisation via the linker (Dr. Xu added a cell-penetrating peptide tag to facilitate studies *in vivo*).

After careful consideration of the crystal structures and following Dr. Xu's recommendations, I opted for two i to i+7 staples between R^{39} and S^{46} for the C-terminal α-helix and between D^{23} and K^{30} for the variable α-helix. Both of these stapling positions correspond to the solvent-exposed face of their respective α-helices, and the chosen residues make no interactions with β-catenin in their helix forms and so these are sterically unhindered positions where the introduction of a staple should have no impact on binding affinity. For these staples I will use azidohomoalanine (aha) as the azide-containing, unnatural amino acid component and 3,5-diethynylbenzoic acid as the dialkyne containing linker.

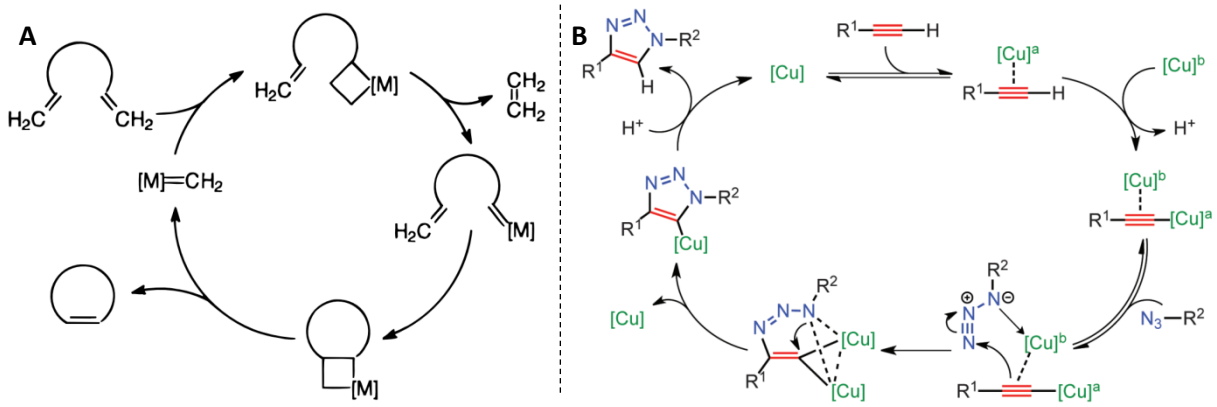


Figure 7-4 – The mechanisms of peptide stapling reactions. (A) and (B) show the reaction mechanisms for ring-closing metathesis (RCM) and copper-catalysed azide-alkyne cycloaddition (CuAAC) reactions respectively. RCM is a common method of stapling peptides, whereas the Itzhaki lab favours CuAAC. (A) and (B) were recreated from Fürstner and Langemann (1996) and Worrell, Malik and Fokin (2013) respectively.

7.1.2 Incorporation of Unnatural Amino Acids using Methionine Auxotrophs

In order to facilitate peptide stapling, I used a method of incorporating the unnatural amino acid similar to that used by Plückthun's group for the bioconjugation of an ANK repeat protein (Simon, Zangemeister-Wittke and Plückthun, 2012). This method takes advantage of the fact that the *E. coli* methionyl-tRNA synthetase is promiscuous and can substitute a number of structural analogues for methionine, including the click-useful amino acids: azidoalanine, azidohomoalanine and 2-amino-5-hexynoic acid (Kiick *et al.*, 2002). When combined with a methionine auxotrophic strain of *E. coli* (I used B834(DE3) cells, see Section 2.1.1), proteins can be expressed containing the desired unnatural amino acid in place of methionine residues.

In order to utilise this method I created the M1L D23M K30M S31C mutational variant construct via a series of SDMs (see Appendix A, Section A.3.5), hereafter referred to as the CLICK31 construct. This construct was made in the His-tagged vector (pRSET) instead of the GST-tagged vector (pRSET-GST), as the GST-fusion tag contains a number of structurally significant methionine residues that would disrupt its ability to function as a purification tag as well as unnecessarily consuming the unnatural amino acid.

The CLICK31-pRSET was transformed into B834(DE3) cells, as described in Section 2.2.4, except 250 µl of SOC medium was used instead of 200 µl of 2xYT medium in the outgrowth step. After overnight incubation a single colony was picked and inoculated into a 1:10 started culture (200 ml) comprised of fresh 2xYT medium, supplemented with 100 mg l^{-1} amp and an additional 1% (w/v) glucose, and grown overnight at 37°C in a shaking incubator.

The 1:10 started culture was inoculated into fresh 2xYT medium (2 l), supplemented with 50 mg l^{-1} amp and 1% (w/v) glucose and incubated at 37°C in a shaking incubator until reaching OD_{600} = 0.6 (3-

4 hrs). This procedure ensured there was a sufficient cell mass for protein production. The cells were then harvested as described in Section 2.3 and the supernatant media discarded. The cell pellet was then washed three times by resuspending in ice-cold 1x PBS, harvesting then discarding the supernatant. This was to remove any residual 2xYT media. The cell pellet was then resuspended and divided up in fresh SelenoMet™ medium (2 l) supplemented with 50 mg l^{-1} amp and 50 mg l^{-1} azidohomoalanine hydrochloride (equivalent to 40 mg l^{-1} of the free amino acid), and incubated at 37°C in a shaking incubator for 30 min prior to being induced with 1 mM IPTG. This delay between resuspension and inducement is to restart cell division and to deplete internal stores of methionine whilst minimising catabolic recycling of methionine. The cells were grown at 30°C for a maximum 4 hrs before being harvested and the pellet frozen for later.

Unfortunately, when the cell pellet was lysed and the protein purified in the usual way (as described in Section 2.5), only unusably small amounts were obtained which were then lost on the G75 column. Due to the cost of the unnatural amino acids, as well as constraints on time, this approach was dropped in favour of a new approach: solid-phase peptide synthesis combined with native chemical ligation.

7.1.3 Solid-Phase Peptide Synthesis and Native Chemical Ligations

Solid-phase peptide synthesis (SPPS) was first pioneered in the 1960's (Merrifield, 1963) and has subsequently become the preferred method of abiologically synthesising peptides. In recent years, this previously time consuming process has been automated and commercialised, providing and alternative method of producing peptides and protein fragments longer than ten residues. However, even with the aid of automation, constructs longer than ≈ 35 residues are difficult to synthesise due to problems with the coupling reaction not proceeding to 100% completion, reducing the overall efficiency. For my 56 residue TCF7L2 constructs this presents a problem, but one that could be easily overcome using native chemical ligation.

Native chemical ligation (NCL) is chemo- and regio-selective reaction for joining two unprotected peptide fragments together into a single polypeptide chain (Dawson *et al.*, 1994), fragments that can be made using SPPS. This versatile technique has been used in the past to synthesise whole proteins from fragments containing unnatural amino acids modifications (Qi *et al.*, 2015) or other natural modifications like glycosylation (Wang *et al.*, 2013).

The reaction requires that the N-terminal peptide (peptide 1) to have its C-terminus modified with a thioester, a process that can be achieved easily when cleaving from the resin if the fragment was made using SPPS, and for the C-terminal peptide (peptide 2) to have a cysteine as its N-terminal-

most residue. This cysteine residue undergoes a reversible transthioesterification reaction with the C-terminal thioester followed by an irreversible S to N acyl shift to the N-terminus of peptide 2 (see Figure 7-5). Because the initial transthioesterification reaction is fully reversible, any cross reactions with other cysteine residues, or thiol containing modifications, cannot proceed and so reverse until the desired reaction and irreversible acyl shift occur, yielding only the desired product.

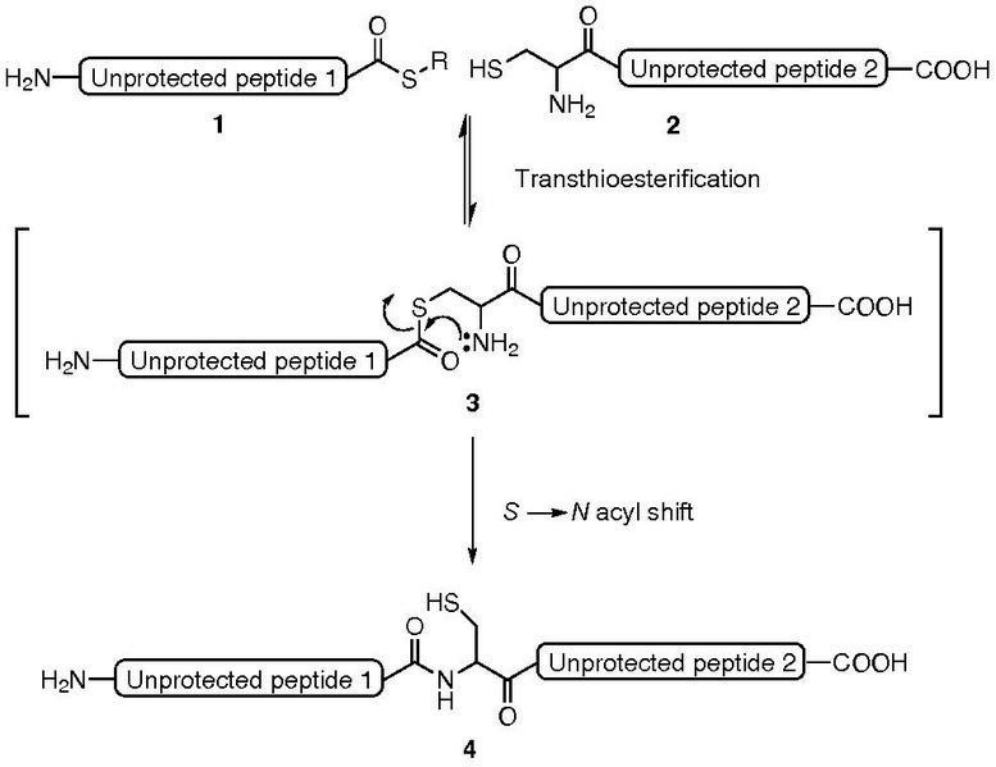


Figure 7-5 – The mechanism of native chemical ligation. Image provided by Wikimedia commons.

By combining these two techniques I was able to adopt a modular approach and would be able to create TCF7L2 constructs that are click-able at the C-terminal α-helix, the variable α-helix or both. I designed four peptides to be synthesised by SPPS: click-able (1_c) and wild-type (1_{wt}) versions of residues 1-30 (including an N-terminal gly-ser- for consistency with my previous findings); and click-able (2_c) and wild-type (2_{wt}) versions of residues 31-54 with the S31C mutation, enabling both NCL and fluorescein maleimide labelling (see Appendix A, Section A.3.5). 1_c and 2_{wt} could be combined to create a construct that was staple-able across the variable α-helix, 1_{wt} and 2_c to make one staple-able across the C-terminal α-helix, and 1_c and 2_c to make the double stapled construct. NCL is permissive enough that it can be performed on peptides that have already been macrocyclised, which would allow the creation of the double stapled construct without fear of cross reactions.

However due to time constraints I was unable to synthesise these constructs. With the C-terminal α-helix, if this helix is fully formed prior to TCF7L2-β-catenin binding then I would expect there to be little change in the association and dissociation profiles. However if it is not pre-formed, I would

expect an increase in the association rate constant as a number of key residues are now pre-positioned to more easily form complex contacts. I would also predict a decrease in dissociation rate constants as the conformational flexibility of this region would be reduced and, in order to dissociate, the contacts of this region would need to be broken simultaneously rather than sequentially. It might also be the case that one of the observed dissociation phases is due to the unravelling of the C-terminal α-helix. Stapling this α-helix would restrict this unravelling, in which case I would predict one of the dissociation phases to no longer be observed. Predicting the effect of stapling the variable α-helix is more difficult given that what I found was that entire region seemed to contribute very little to binding (see Chapter 5). But I would predict that what would happen is a decrease in both association and dissociation rates, similar to those seen for the XXXXX S31C construct.

I would still need to test the effect of the staple on α-helical propensities of the stapled regions. α-helical structure can be measured by circular dichroism (CD) spectroscopy, and is characterised by a local minima in ellipticity at 222 nm (θ_{222}) and can also be induced by adding TFE to the solution (Jasanoff and Fersht, 1994).

I did test the α-helical propensity of the wild-type construct (TCF7L2 (1-54)) during the method devolopment and found that this construct showed a very small amount of α-helical structure at 0% TFE which increased to an apparent plateau above 45% TFE (see Figure 7-6). At the time, further CD-based experiments were disregarded as any CD-signal changes that might be generated by TCF7L2 would likely be masked by the signal from β-catenin (which contains 36 long α-helices).

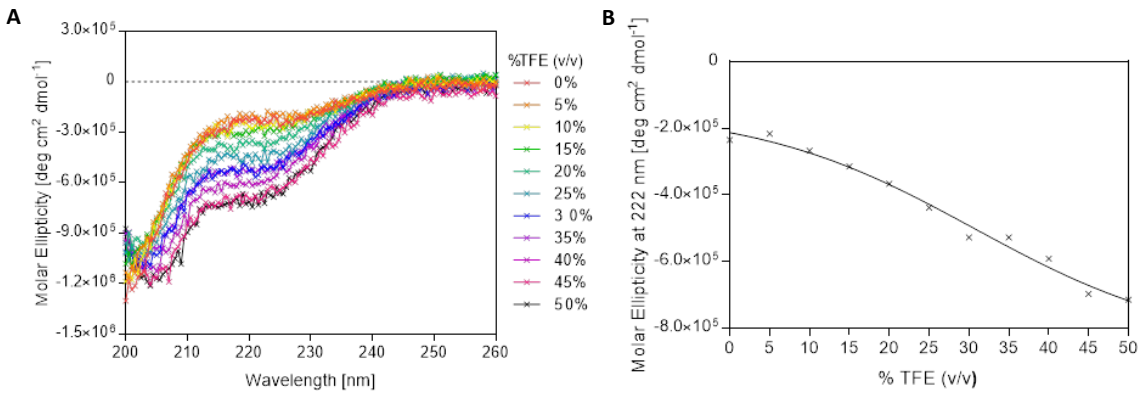


Figure 7-6 – α-helicity test of the TCF7L2 (1-54) construct. (A) shows how the CD spectra between 200 and 260 nm changes with increasing TFE concentration and (B) highlights how the minima at 222 nm changes with TFE concentration. Absorbance at 222 nm is known to be linear with α-helix content and (B) shows the expected sigmoidal curve as % helical content increases from 0 to 50% TFE. CD spectra were recorded on a Chirascan™ spectropolarimeter (Applied Photophysics) using a cell of pathlength 1 mm. This experiment was performed in 25 mM potassium phosphate pH 6.0, 75 mM NaCl and 1 mM DTT with between 5 and 50% (v/v) TFE at 25°C. Final concentration of TCF7L2 (1-54) was 10.3 μM.

However changes in α-helical propensity on stapling could be determined by performing similar experiments on the 1_c and 2_c NCL fragments before and after a successful click reaction. If the staple has the expected effect on α-helical propensity, I would expect to see a smaller magnitude change in θ_{222}, reaching a similar final θ_{222}, after stapling than before stapling, and if the stapling has no effect then I would expect the pre-stapling spectra and the post-stapling spectra to be identical. If it is the case that the helix is preformed, as may be the case for the C-terminal helix (2_c), then I would expect there to be little to no change on addition of TFE and the before stapling spectra and the after stapling spectra to be identical.

7.1.4 Truncated Constructs, Tethered Constructs and Avidity

In order to investigate the proposed two-site avidity model I designed a series truncated TCF7L2 constructs representing these separate regions to test, with the intension of subsequently re-attaching the two domains with various other unnatural likers, e.g. various lengths of PEG or polyalanine, to determine the contribution of the variable region to binding beyond simply being a tether.

One of these constructs, the "WT" S31C (1-32) construct, was made by RTH PCR in both the pRSET and pRSET-GST vectors. However the his-tagged construct did not express well, and the GST-tagged construct did not cleave efficiently with thrombin. Instead the "WT" S31C (1-31) construct was successfully made using SPPS, which also served as a practice run at using the peptide synthesiser before using unnatural amino acids. However, again due to time constraints, this construct was never tested.

In Omer *et al.*'s original paper on TCF7L2-β-catenin binding (Omer *et al.*, 1999), they also studied a range of truncated constructs of TCF7L2. They tested a TCF7L2 construct composed of residues 1-30 which had a K_d 400 times higher than a construct composed of residues 1-53 (6 ± 2.4 μM compared to 15 ± 6 nM), which is comparable to my wild-type TCF7L2 (1-54) construct. They also tested TCF7L2 constructs composed of residues 38-54, 30-100 and 20-100 and saw no significant binding from any of these constructs, although a 12-100 construct did have a K_d of 85 ± 8 nM, 6 times that of their 1-53 construct. They also tested a 1-42 construct with a K_d of 9 ± 1.1 μM, not significantly different from their 1-30 construct. Taken together, their data further support the idea that the variable region contributes little to binding and is merely acting as a tether between binding regions.

7.2 Future Work

The biggest issue during this study was the lack of a reliable method for determining the equilibrium dissociation constant, $K_{d,eq}$. While my $K_{d,eq}$ (28 nM), as determined by ITC, was consistent with previously reported $K_{d,eq}$s (15 ± 3 nM), the calculated kinetic dissociation constant, $K_{d,kin}$, was three orders of magnitudes lower than expected (7.8 ± 0.6 pM). When I tried to determine $K_{d,eq}$ using an alternative method, i.e. fluorescence anisotropy titration, I was unable to produce a consistent value (see Section 4.6). However when we compare the $K_{d,eq}$s for wild-type constructs from different alanine scans from the literature, each of which uses a different technique to determine $K_{d,eq}$, we can see that there is a difference of one order of magnitude between each of the data sets (see Table 7-1). This inconsistency suggests that this issue is more complex than it first appears.

A more reliable method of determining $K_{d,eq}$ is required. I have previously suggested in Section 4.8 that the incredibly long dissociation times of the TCF7L2-β-catenin complex may be responsible for the inconsistency between my own and literature equilibrium binding data, and my kinetic binding data. The simplest and likely most reliable way of determining $K_{d,eq}$ which can account for the long dissociation time would be a plate-based direct or competition fluorescence polarisation assay as the plate can be prepared then incubated for an appropriate amount of time prior to reading.

A direct fluorescence polarisation assay can be achieved using a serial dilution of β-catenin to which a fixed known concentration of labelled-TCF7L2 is added then incubated before reading. The $K_{d,eq}$ can be calculated from this assay using the polarisation, accounting for ligand depletion equation (Equation 2-8). Competition assays would involve adding a fixed known concentration of pre-complexed labelled-TCF7L2 and β-catenin to a serial dilution of the competitor variant that was being tested. These data can be fit with the equations described in Wang (1995) to determine the $K_{d,eq}$ of the competitor variant.

These two methods have been successfully used by other members of the Itzhaki Lab in the past and have been used by Dr. Dunbar to study the interaction between a stapled axin construct and β-catenin. I had previously considered using these methods however I decided to use polarisation titration instead as plate-based assays used large amounts of β-catenin. Dr. Dunbar herself only performed a few such experiments.

Further research into the multi-phasic dissociation is also necessary. As I mentioned in Section 6.4, the only construct for which I observed four dissociation phases was the I19A S31C construct. This was because the dissociation reaction was such that it could be measured using both stopped-flow and time-dependent fluorescence spectroscopy. However, F21A S31C also had dissociation kinetics

that allowed it to be measured using either technique, and it did have a measurable third dissociation phase. Additionally the E17A S31C variant also had suitable dissociation kinetics as to be measure by either technique. This would be the starting point, but three mutations are not enough to strongly support the theory so I would also want to try and test the other residues in that region that I did not test (D^{16}, L^{18}, S^{20} and K^{22}) to see if a similar pattern could be seen for any of them.

The variable region also warrants further characterisation, particularly residues 23-29. Mutations in this region had a negligible impact on complex association and dissociation at physiological salt concentrations, however at higher concentrations, they had a much more dramatic effect on the kinetics. This suggests a fuzzy mode of binding for this region. Fuzzy complexes can be characterised by a number of techniques, primarily NMR, SAXS, and HDX-MS (Sibille and Bernadó, 2012; Balasubramaniam and Komives, 2013) of which HDX-MS would be simplest way to investigate this system. NMR and SAXS are best used in combination with each other and would be able to provide more quantitative data than HDX-MS but are more time-consuming and expensive. HDX-MS would be able to give a qualitative measurement which would tell us whether NMR and SAXS were worth pursuing.

In brief; β-catenin would be incubated in D_2O for an amount of time such that the solvent exposed surface hydrogens can exchange with deuterium, but not buried ones. Then TCF7L2 is added before being diluted in H_2O to exchange the deuterium back to hydrogen. This reaction is quenched at various time points, then digested by protease and the fragments analysed by MS, typically MALDI-TOF. From this, the rate of exchange for hydrogens in different positions on β-catenin, in the presence or absence of TCF7L2, can be determined. Hydrogens that are shielded from the solvent by the presence of TCF7L2 should exchange slower as the TCF7L2 would need to dissociate before they could exchange and, from this a "protection factor" can be calculated (Balasubramaniam and Komives, 2013).

I would expect the variable region of TCF7L2 to provide less shielding (lower protection factor), if any, to the β-catenin hydrogens than the other regions. Assuming there is some protection, the effect of osmolarity on this region can be investigated further, as the hydrogen/deuterium exchange steps can be performed at various salt concentrations. This approach could potentially also be used to investigate the effect of the alanine scan mutational variants on binding, to see whether these mutations disrupt binding more locally or more globally across the TCF7L2- β-catenin interface.

Another aspect that merits investigation is what effect labelling at position 46 has on the observed binding kinetics. As I mentioned in Section 4.6, the "WT" S46C and "WT" S20C were created to acts

as local reporters of the C- and N-terminal regions respectively. While position 20 was shown to be unsuitable for reporting the binding interaction, it would be interesting to see if the two phase/four phase pattern in the dissociation kinetics was observed when TCF7L2 is labelled at S46C. For most of the alanine variants used in Chapters 5 and 6, I also created plasmids containing the equivalent S46C construct, but did not express or purify these constructs due to time constraints.

In summary there is a lot more still to be learnt about the TCF7L2-β-catenin interaction. Future work on this project needs to focus on the development of a more reliable method of determining the equilibrium dissociation constant in order to further quantify the kinetics of the binding interaction.

Appendix A – DNA and Proteins Sequences

A.1 LRRK2

Proposed repeating regions as predicted by (Mills *et al.*, 2012):

- ARM Domain
- ANK Domain
- LRR Domain
- ROC Domain
- Kinase Domain
- WD40 Domain

```
        10         20         30         40         50         60         70
MASGSCQGCE EDEETLKKLI VRLNNVQEGK QIETLVQILE DLLVFTYSER ASKLFQGKNI HVPLLIVLDS
        80         90        100        110        120        130        140
YMRVASVQQV GWSLLCKLIE VCPGTMQSLM GPQDVGNDWE VLGVHQLILK MLTVHNASVN LSVIGLKTLD
       150        160        170        180        190        200        210
LLLTSGKITL LILDEESDIF MLIFDAMHSF PANDEVQKLG CKALHVLFER VSEEQLTEFV ENKDYMILLS
       220        230        240        250        260        270        280
ALTNFKDEEE IVLHVLHCLH SLAIPCNNVE VLMSGNVRCY NIVVEAMKAF PMSERIQEVS CCLLHRLTLG
       290        300        310        320        330        340        350
NFFNILVLNE VHEFVVKAVQ QYPENAALQI SALSCLALLT ETIFLNQDLE EKNENQENDD EGEEDKLFWL
       360        370        380        390        400        410        420
EACYKALTWH RKNKHVQEAA CWALNNLLMY QNSLHEKIGD EDGHFPAHRE VMLSMLMHSS SKEVFQASAN
       430        440        450        460        470        480        490
ALSTLLEQNV NFRKILLSKG IHLNVLELMQ KHIHSPEVAE SGCKMLNHLF EGSNTSLDIM AAVVPKILTV
       500        510        520        530        540        550        560
MKRHETSLPV QLEALRAILH FIVPGMPEES REDTEFHHKL NMVKKQCFKN DIHKLVLAAL NRFIGNPGIQ
       570        580        590        600        610        620        630
KCGLKVISSI VHFPDALEML SLEGAMDSVL HTLQMYPDDQ EIQCLGLSLI GYLITKKNVF IGTGHLLAKI
       640        650        660        670        680        690        700
LVSSLYRFKD VAEIQTKGFQ TILAILKLSA SFSKLLVHHS FDLVIFHQMS SNIMEQKDQQ FLNLCCKCFA
       710        720        730        740        750        760        770
KVAMDDYLKN VMLERACDQN NSIMVECLLL LGADANQAKE GSSLICQVCE KESSPKLVEL LLNSGSREQD
       780        790        800        810        820        830        840
VRKALTISIG KGDSQIISLL LRRLALDVAN NSICLGGFCI GKVEPSWLGP LFPDKTSNLR KQTNIASTLA
       850        860        870        880        890        900        910
RMVIRYQMKS AVEEGTASGS DGNFSEDVLS KFDEWTFIPD SSMDSVFAQS DDLDSEGSEG SFLVKKKSNS
       920        930        940        950        960        970        980
ISVGEFYRDA VLQRCSPNLQ RHSNSLGPIF DHEDLLKRKR KILSSDDSLR SSKLQSHMRH SDSISSLASE
       990       1000       1010       1020       1030       1040       1050
REYITSLDLS ANELRDIDAL SQKCCISVHL EHLEKLELHQ NALTSFPQQL CETLKSLTHL DLHSNKFTSF
      1060       1070       1080       1090       1100       1110       1120
PSYLLKMSCI ANLDVSRNDI GPSVVLDPTV KCPTLKQFNL SYNQLSFVPE NLTDVVEKLE QLILEGNKIS
      1130       1140       1150       1160       1170       1180       1190
GICSPLRLKE LKILNLSKNH ISSLSENFLE ACPKVESFSA RMNFLAAMPF LPPSMTILKL SQNKFSCIPE
      1200       1210       1220       1230       1240       1250       1260
AILNLPHLRS LDMSSNDIQY LPGPAHWKSL NLRELLFSHN QISILDLSEK AYLWSRVEKL HLSHNKLKEI
      1270       1280       1290       1300       1310       1320       1330
PPEIGCLENL TSLDVSYNLE LRSFPNEMGK LSKIWDLPLD ELHLNFDFKH IGCKAKDIIR FLQQRLKKAV
      1340       1350       1360       1370       1380       1390       1400
PYNRMKLMIV GNTGSGKTTL LQQLMKTKKS DLGMQSATVG IDVKDWPIQI RDKRKRDLVL NVWDFAGREE
      1410       1420       1430       1440       1450       1460       1470
FYSTHPHFMT QRALYLAVYD LSKGQAEVDA MKPWLFNIKA RASSSPVILV GTHLDVSDEK QRKACMSKIT
      1480       1490       1500       1510       1520       1530       1540
KELLNKRGFP AIRDYHFVNA TEESDALAKL RKTIINESLN FKIRDQLVVG QLIPDCYVEL EKIILSERKN
      1550       1560       1570       1580       1590       1600       1610
VPIEFPVIDR KRLLQLVREN QLQLDENELP HAVHFLNESG VLLHFQDPAL QLSDLYFVEP KWLCKIMAQI
      1620       1630       1640       1650       1660       1670       1680
LTVKVEGCPK HPKGIISRRD VEKFLSKKRK FPKNYMSQYF KLLEKFQIAL PIGEEYLLVP SSLSDHRPVI
```

```
      1690       1700       1710       1720       1730       1740       1750
ELPHCENSEI IIRLYEMPYF PMGFWSRLIN RLLEISPYML SGRERALRPN RMYWRQGIYL NWSPEAYCLV
      1760       1770       1780       1790       1800       1810       1820
GSEVLDNHPE SFLKITVPSC RKGCILLGQV VDHIDSLMEE WFPGLLEIDI CGEGETLLKK WALYSFNDGE
      1830       1840       1850       1860       1870       1880       1890
EHQKILLDDL MKKAEEGDLL VNPDQPRLTI PISQIAPDLI LADLPRNIML NNDELEFEQA PEFLLGDGSF
      1900       1910       1920       1930       1940       1950       1960
GSVYRAAYEG EEVAVKIFNK HTSLRLLRQE LVVLCHLHHP SLISLLAAGI RPRMLVMELA SKGSLDRLLQ
      1970       1980       1990       2000       2010       2020       2030
QDKASLTRTL QHRIALHVAD GLRYLHSAMI IYRDLKPHNV LLFTLYPNAA IIAKIADYGI AQYCCRMGIK
      2040       2050       2060       2070       2080       2090       2100
TSEGTPGFRA PEVARGNVIY NQQADVYSFG LLLYDILTTG GRIVEGLKFP NEFDELEIQG KLPDPVKEYG
      2110       2120       2130       2140       2150       2160       2170
CAPWPMVEKL IKQCLKENPQ ERPTSAQVFD ILNSAELVCL TRRILLPKNV IVECMVATHH NSRNASIWLG
      2180       2190       2200       2210       2220       2230       2240
CGHTDRGQLS FLDLNTEGYT SEEVADSRIL CLALVHLPVE KESWIVSGTQ SGTLLVINTE DGKKRHTLEK
      2250       2260       2270       2280       2290       2300       2310
MTDSVTCLYC NSFSKQSKQK NFLLVGTADG KLAIFEDKTV KLKGAAPLKI LNIGNVSTPL MCLSESTNST
      2320       2330       2340       2350       2360       2370       2380
ERNVMWGGCG TKIFSFSNDF TIQKLIETRT SQLFSYAAFS DSNIITVVVD TALYIAKQNS PVVEVWDKKT
      2390       2400       2410       2420       2430       2440       2450
EKLCGLIDCV HFLREVMVKE NKESKHKMSY SGRVKTLCLQ KNTALWIGTG GGHILLLDLS TRRLIRVIYN
      2460       2470       2480       2490       2500       2510       2520
FCNSVRVMMT AQLGSLKNVM LVLGYNRKNT EGTQKQKEIQ SCLTVWDINL PHEVQNLEKH IEVRKELAEK
   2527
MRRTSVE
```

All constructs include a N-terminal GS due to the thrombin cleavage site

A.1.1 ARM Construct (1-661)

```
GSMASGSCQGCEEDEETLKKLIVRLNNVQEGKQIETLVQILEDLLVFTYSERASKLFQGKNIHVPLLIVLDSYMRVASVQQVG
WSLLCKLIEVCPGTMQSLMGPQDVGNDWEVLGVHQLILKMLTVHNASVNLSVIGLKTLDLLLTSGKITLLILDEESDIFMLIF
DAMHSFPANDEVQKLGCKALHVLFERVSEEQLTEFVENKDYMILLSALTNFKDEEEIVLHVLHCLHSLAIPCNNVEVLMSGNV
RCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNILVLNEVHEFVVKAVQQYPENAALQISALSCLALLTETIFLNQDLE
EKNENQENDDEGEEDKLFWLEACYKALTWHRKNKHVQEAACWALNNLLMYQNSLHEKIGDEDGHFPAHREVMLSMLMHSSSKE
VFQASANALSTLLEQNVNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFEGSNTSLDIMAAVVPKILTVMKRHET
SLPVQLEALRAILHFIVPGMPEESREDTEFHHKLNMVKKQCFKNDIHKLVLAALNRFIGNPGIQKCGLKVISSIVHFPDALEM
LSLEGAMDSVLHTLQMYPDDQEIQCLGLSLIGYLITKKNVFIGTGHLLAKILVSSLYRFKDVAEIQTKGFQTILAILKLSAS
```

A.1.2 ANK Construct (668-908)

```
GSMHHSFDLVIFHQMSSNIMEQKDQQFLNLCCKCFAKVAMDDYLKNVMLERACDQNNSIMVECLLLLGADANQAKEGSSLICQ
VCEKESSPKLVELLLNSGSREQDVRKALTISIGKGDSQIISLLLRRLALDVANNSICLGGFCIGKVEPSWLGPLFPDKTSNLR
KQTNIASTLARMVIRYQMKSAVEEGTASGSDGNFSEDVLSKFDEWTFIPDSSMDSVFAQSDDLDSEGSEGSFLVKKKS
```

A.1.3 LRR Construct (979-1328)

```
GSMSEREYITSLDLSANELRDIDALSQKCCISVHLEHLEKLELHQNALTSFPQQLCETLKSLTHLDLHSNKFTSFPSYLLKMS
CIANLDVSRNDIGPSVVLDPTVKCPTLKQFNLSYNQLSFVPENLTDVVEKLEQLILEGNKISGICSPLRLKELKILNLSKNHI
SSLSENFLEACPKVESFSARMNFLAAMPFLPPSMTILKLSQNKFSCIPEAILNLPHLRSLDMSSNDIQYLPGPAHWKSLNLRE
LLFSHNQISILDLSEKAYLWSRVEKLHLSHNKLKEIPPEIGCLENLTSLDVSYNLELRSFPNEMGKLSKIWDLPLDELHLNFD
FKHIGCKAKDIIRFLQQRLKK
```

A.1.4 WD40 Construct (2159-2527)

```
GSMHHNSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESWIVSGTQSGTLLVINTEDGKKRHTL
EKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGT
KIFSFSNDFTIQKLIETRTSQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLREVMVKENKES
KHKMSYSGRVKTLCLQKNTALWIGTGGGHILLLDLSTRRLIRVIYNFCNSVRVMMTAQLGSLKNVMLVLGYNRKNTEGTQKQK
EIQSCLTVWDINLPHEVQNLEKHIEVRKELAEKMRRTSVE
```

A.1.5 Extended ARM Construct

GSMASGSCQGCEEDEETLKKLIVRLNNVQEGKQIETLVQILEDLLVFTYSERASKLFQGKNIHVPLLIVLDSYMRVASVQQVG
WSLLCKLIEVCPGTMQSLMGPQDVGNDWEVLGVHQLILKMLTVHNASVNLSVIGLKTLDLLLTSGKITLLILDEESDIFMLIF
DAMHSFPANDEVQKLGCKALHVLFERVSEEQLTEFVENKDYMILLSALTNFKDEEEIVLHVLHCLHSLAIPCNNVEVLMSGNV
RCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNILVLNEVHEFVVKAVQQYPENAALQISALSCLALLTETIFLNQDLE
EKNENQENDDEGEEDKLFWLEACYKALTWHRKNKHVQEAACWALNNLLMYQNSLHEKIGDEDGHFPAHREVMLSMLMHSSSKE
VFQASANALSTLLEQNVNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFEGSNTSLDIMAAVVPKILTVMKRHET
SLPVQLEALRAILHFIVPGMPEESREDTEFHHKLNMVKKQCFKNDIHKLVLAALNRFIGNPGIQKCGLKVISSIVHFPDALEM
LSLEGAMDSVLHTLQMYPDDQEIQCLGLSLIGYLITKKNVFIGTGHLLAKILVSSLYRFKDVAEIQTKGFQTILAILKLSASK
LLVHHSFDLVIFHQMSSNIMEQKDQQFLNLCCKCFAK

A.2 Weis β-catenin (134-671)

GSHAVVNLINYQDDAELATRAIPELTKLLNDEDQVVVNKAAVMVHQLSKKEASRHAIMRSPQMVSAIVRTMQNTNDVETARCT
AGTLHNLSHHREGLLAIFKSGGIPALVKMLGSPVDSVLFYAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAIT
TDCLQILAYGNQESKLIILASGGPQALVNIMRTYTYEKLLWTTSRVLKVLSVCSSNKPAIVEAGGMQALGLHLTDPSQRLVQN
CLWTLRNLSDAATKQEGMEGLLGTLVQLLGSDDINVVTCAAGILSNLTCNNYKNKMMVCQVGGIEALVRTVLRAGDREDITEP
AICALRHLTSRHQEAEMAQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAIPRLVQLLVRAHQD
TQRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNRIVIRGLNTIPLFVQLLYSPIENIQRVAAGVLCELAQDKEAAE
AIEAEGATAPLTELLHSRNEGVATYAAAVLFRMSEDKPQDYK

A.3 TCF7L2

All TCF7L2 constructs expressed in *E. coli* were expressed from vectors containing N-terminal, thrombin-cleavable purification tags. The process of cleaving the tag with thrombin leaves behind an N-terminal GS- at the beginning of the protein sequence. The constructs produced by peptide synthesis had an additional GS- added to the N-terminus of the sequence so their findings would be consistent with those obtained for constructs expressed in *E. coli*.

A.3.1 Wild-Type TCF7L2 (1-54)

GSMPQLNGGGGDDLGANDELISFKDEGEQEEKSSENSSAERDLADVKSSLVNESET

A.3.2 Pseudo Wild-Type Constructs

"WT" S31C

GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCSENSSAERDLADVKSSLVNESET

"WT" S20C

GSMPQLNGGGGDDLGANDELICFKDEGEQEEKSSENSSAERDLADVKSSLVNESET

"WT" S46C

GSMPQLNGGGGDDLGANDELISFKDEGEQEEKSSENSSAERDLADVKCSLVNESET

A.3.3 Ensemble Binding Constructs

E24A S31C

GSMPQLNGGGGDDLGANDELISFKDAGEQEEKCSENSSAERDLADVKSSLVNESET

E29A S31C

GSMPQLNGGGGDDLGANDELISFKDEGEQEAKCSENSSAERDLADVKSSLVNESET

E24A E29A S31C

GSMPQLNGGGGDDLGANDELISFKDAGEQEAKCSENSSAERDLADVKSSLVNESET

D23A S31C
GSMPQLNGGGGDDLGANDELISFKAEGEQEEKCSENSSAERDLADVKSSLVNESET

E26A S31C
GSMPQLNGGGGDDLGANDELISFKDEGAQEEKCSENSSAERDLADVKSSLVNESET

E28A S31C
GSMPQLNGGGGDDLGANDELISFKDEGEQAEKCSENSSAERDLADVKSSLVNESET

XXXXX S31C (D23A E24A E26A E28A E29A S31C)
GSMPQLNGGGGDDLGANDELISFKAAGAQAAKCSENSSAERDLADVKSSLVNESET

A.3.4 Alanine Scan Constructs

L12A S31C
GSMPQLNGGGGDDAGANDELISFKDEGEQEEKCSENSSAERDLADVKSSLVNESET

N15A S31C
GSMPQLNGGGGDDLGAADELISFKDEGEQEEKCSENSSAERDLADVKSSLVNESET

E17A S31C
GSMPQLNGGGGDDLGANDALISFKDEGEQEEKCSENSSAERDLADVKSSLVNESET

I19A S31C
GSMPQLNGGGGDDLGANDELASFKDEGEQEEKCSENSSAERDLADVKSSLVNESET

F21A S31C
GSMPQLNGGGGDDLGANDELISAKDEGEQEEKCSENSSAERDLADVKSSLVNESET

E38A S31C
GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCSENSSAARDLADVKSSLVNESET

L41A S31C
GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCSENSSAERDAADVKSSLVNESET

V44A S31C
GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCSENSSAERDLADAKSSLVNESET

K45A S31C
GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCSENSSAERDLADVASSLVNESET

L48A S31C
GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCSENSSAERDLADVKSSAVNESET

A.3.5 Staple-able Constructs

CLICK31 (M1L D23M K30M S31C)
GSLPQLNGGGGDDLGANDELISFKMEGEQEEMCSENSSAERDLADVKSSLVNESET

Native Chemical Ligation Fragments
1_{wt} – GSMPQLNGGGGDDLGANDELISFKDEGEQEEK
1_{c} – GSMPQLNGGGGDDLGANDELISFKXEGEQEEX
2_{wt} – CSENSSAERDLADVKSSLVNESET
2_{c} – CSENSSAEXDLADVKXSLVNESET
X = azidohomoalanine

A.3.6 Truncated TCF7L2 Constructs

"WT" S31C (1-32)

GSMPQLNGGGGDDLGANDELISFKDEGEQEEKCS

This construct was made accidentally due to poor primer design. The intended construct was the "WT" S31C (1-31) construct described below.

"WT" S31C (1-31)

GSMPQLNGGGGDDLGANDELISFKDEGEQEEKC

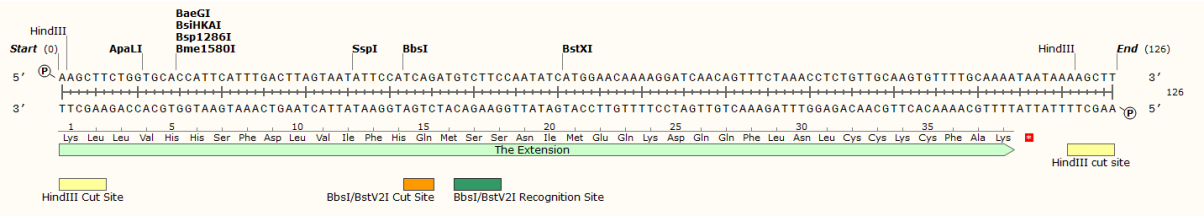
"WT" S31C (31-54)

CSENSSAERDLADVKSSLVNESET

Appendix B – DNA Primers and Gene Strings

B.1 LRRK2

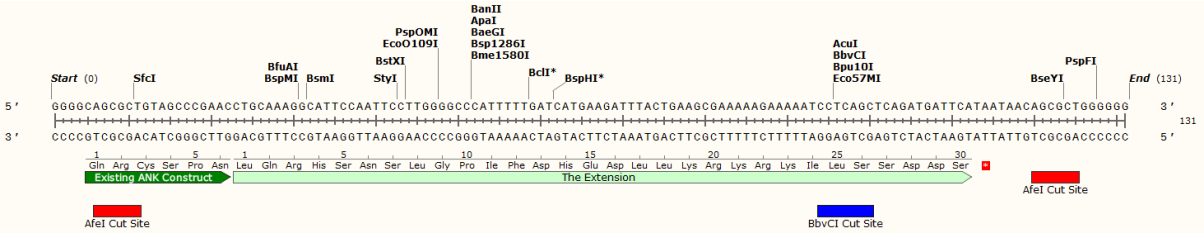
B.1.1 C-terminal ARM Construct Extension



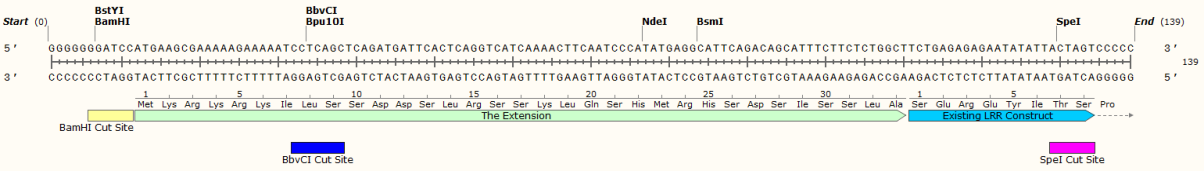
B.1.2 ARM Extension Extending Primers

Forward: 5′-GGGGCAAGCTTCTGGTGCACCATTCATTTG-3′
Reverse: 5′-CCCCCAAGCTTTTATTATTTTGCAAAACAC-3′

B.1.3 ANK C-terminal Extension



B.1.4 LRR N-terminal Extension



B.2 TCF7L2

B.2.1 Pseudo Wild-Type SDM Primers

“WT” S31C

Forward 5′-GCGAGCAAGAAGAAAAATGTAGCGAAAATAGC-3′
Reverse 5′-CTGCGCTGCTATTTTCGCTACATTTTTCTTCTTG-3′

“WT” S20C

Forward 5′-AACTGATCTGCTTTAAAGATGAA-3′
Reverse 5′-TTTAAAGCAGATCAGTTCATCAT-3′

“WT” S46C

Forward 5′-GATGTTAAATGCAGCCTGG-3′
Reverse 5′-CCAGGCTGCATTTAACATC-3′

B.2.2 Ensemble Binding SDM Primers

E24A S31C (RTH)

Forward 5′-CAGGCGAGCAAGAAGAAAAAAG-3′

Reverse 5′-CATCTTTAAAGCTGATCAGTTCATCA-3′

E29A S31C (RTH)

Forward 5′-CAAAAAGCAGCGAAAATAGCAGC-3′
Reverse 5′-CTTCTTGCTCGCCTTCATCTTTA-3′

E29A S31C → E24A E29A S31C

Forward 5′-GCTTTAAAGATGCAGGCGAGCAAGAAG-3′
Reverse 5′-CTCGCCTGCATCTTTAAAGCTGATCAG-3′

D23A S31C

Forward 5′-CAGCTTTAAAGCAGAAGGCGAGCAAGAAG-3′
Reverse 5′-TCGCCTTCTGCTTTAAAGCTGATCAGTTCATC-3′

E26A S31C

Forward 5′-GGCGCGCAAGAAGAAAAATGTAG-3′
Reverse 5′-TTGCGCGCCTTCATCTTTAAAGC-3′

E28A S31C

Forward 5′-GAAGGCGAGCAAGCAGAAAAATG-3′
Reverse 5′-GCTACATTTTTCTGCTTGCTCGC-3′

XXXXX S31C (D23A E24A E26A E28A E29A S31C) (RTH)

Forward 5′-GCGCAAGCAGCAAAATGTAG-3′
Reverse 5′-GCCTGCTGCTTTAAAGCTGAT-3′

OR

Forward 5′-CAAGCAGCAAAATGTAGCG-3′
Reverse 5′-TGCGCCTGCTGCTTTAAAG-3′

Both pairs worked

B.2.3 Alanine Scan Constructs SDM Primers

L12A S31C

Forward 5′-GGTGATGATGCGGGTGCAAATGATGAACTGATCAGC-3′
Reverse 5′-GCACCCGCATCATCACCGCCACCACC-3′

N15A S31C (RTH)

Forward 5′-GCGGATGAACTGATCAGCTTTAAAGATG-3′
Reverse 5′-TGCACCCAGATCATCACC-3′

E17A S31C (RTH)

Forward 5′-CGCTGATCAGCTTTAAAGATGAAGG-3′
Reverse 5′-CATCATTTGCACCCAGATC-3′

I19A S31C

Forward 5′-GATGAACTGGCCAGCTTTAAAGATGAAGGCGAGCAAG-3′
Reverse 5′-CTTTAAAGCTGGCCAGTTCATCATTTGCACCCAGATC-3′

F21A S31C (RTH)

Forward 5′-GCGAAAGATGAAGGCGAGCAAG-3′
Reverse 5′-GCTGATCAGTTCATCATTTGC-3′

E38A S31C (RTH)

Forward 5′-CGCGTGATCTGGCAGATGTTAA-3′
Reverse 5′-CTGCGCTGCTATTTTCG-3′

L41A S31C (RTH)

Forward 5′-GCGGCAGATGTTAAAAGCAGCC-3′
Reverse 5′-ATCACGTTCTGCGCTGC-3′

V44A S31C

Forward 5′-CTGGCAGATGCTAAAAGCAGCCTGG-3′
Reverse 5′-CTTTTAGCATCTGCCAGATCACGTTCTGC-3′

K45A S31C (RTH)

Forward 5′-GCGAGCAGCCTGGTTAATGAAA-3′
Reverse 5′-AACATCTGCCAGATCACGTT-3′

L48A S31C

Forward 5′-AAAAGCAGCGCGGTTAATGAAAGCGAAACCTG-3′
Reverse 5′-GCTTTCATTAACCGCGCTGCTTTTAACATCTGCCAG-3′

B.3.4 Staple-able Constructs

M1L S31C

Forward 5′-GGATCCCTGCCGCAGCTGAATG-3′
Reverse 5′-GCGGCAGGGATCCACGC-3′

D23M S31C

Forward 5′-CTGATCAGCTTTAAAATGGAAGGCGAGCAAGAAG-3′
Reverse 5′-GCTCGCCTTCCATTTTAAAGCTGATCAGTTCATC-3′

K30M S31C

Forward 5′-GAAGAAATGTGTAGCGAAAATAGCAGCGCAG-3′
Reverse 5′-CGCTACACATTTCTTCTTGCTCGCC-3′

None of these primer pairs impacts the others so the K1M D23M K30M S31C (CLICK31) construct was made by sequential use of these three paris.

B.3.5 Truncated TCF7L2 Constructs

“WT” S31C (1-32)

Forward 5′-GAAAAATGTAGCTAATAAAGCAGCGCAGAACG-3′
Reverse 5′-GCTGCTTTATTAGCTACATTTTTCTTCTTGCTCG-3′

This construct was made accidentally due to poor primer design. The intended construct was the “WT” S31C (1-31) construct described below.

Bibliography

Abajian, C., Yatsunyk, L. A., Ramirez, B. E. and Rosenzweig, A. C. (2004) 'Yeast Cox17 Solution Structure and Copper(I) Binding.', *Journal of Biological Chemistry*, 279(51), pp. 53584–53592. doi: 10.1074/jbc.M408099200.

Aliaga, L., Cai, H. and Liu, G. (2012) 'α-Synuclein, LRRK2 and their Interplay in Parkinson's Disease.', *Future Neurology*, 7(2), pp. 145–153. doi: 10.2217/fnl.12.2.

Andrade, M. A., Petosa, C., O'Donoghue, S. I., Müller, C. W. and Bork, P. (2001) 'Comparison of ARM and HEAT Protein Repeats.', *Journal of Molecular Biology*, 309(1), pp. 1–18. doi: 10.1006/jmbi.2001.4624.

Anfinsen, C. B. (1973) 'Principles that Govern the Folding of Protein Chains.', *Science*, 181(4096), pp. 223–230. doi: 10.1126/science.181.4096.223.

Arai, M., Ferreon, J. C. and Wright, P. E. (2012) 'Quantitative Analysis of Multisite Protein–Ligand Interactions by NMR: Binding of Intrinsically Disordered p53 Transactivation Subdomains with the TAZ2 Domain of CBP.', *Journal of the American Chemical Society*, 134(8), pp. 3792–3803. doi: 10.1021/ja209936u.

Arce, L., Yokoyama, N. N. and Waterman, M. L. (2006) 'Diversity of LEF/TCF Action in Development and Disease.', *Oncogene*, 25(57), pp. 7492–7504. doi: 10.1038/sj.onc.1210056.

Archbold, H. C., Yang, Y. X., Chen, L. and Cadigan, K. M. (2012) 'How do they do Wnt they do?: Regulation of Transcription by the Wnt/β-catenin Pathway.', *Acta Physiologica*, 204(1), pp. 74–109. doi: 10.1111/j.1748-1716.2011.02293.x.

Arranz, A. M., Delbroek, L., Van Kolen, K., Guimaraes, M. R., Mandemakers, W., Daneels, G., Matta, S., Calafate, S., Shaban, H., Baatsen, P., De Bock, P.-J., Gevaert, K., Vanden Berghe, P., Verstreken, P., De Strooper, B. and Moechars, D. (2015) 'LRRK2 Functions in Synaptic Vesicle Endocytosis through a Kinase-dependent Mechanism.', *Journal of Cell Science*, 128(3), pp. 541–552. doi: 10.1242/jcs.158196.

Aumiller, W. M. and Keating, C. D. (2017) 'Experimental Models for Dynamic Compartmentalization of Biomolecules in Liquid Organelles: Reversible Formation and Partitioning in Aqueous Biphasic Systems.', *Advances in Colloid and Interface Science*, 239, pp. 75–87. doi: 10.1016/j.cis.2016.06.011.

Avalos, J. L., Celic, I., Muhammad, S., Cosgrove, M. S., Boeke, J. D. and Wolberger, C. (2002)

'Structure of a Sir2 Enzyme Bound to an Acetylated p53 Peptide.', *Molecular Cell*, 10(3), pp. 523–535. doi: 10.1016/S1097-2765(02)00628-7.

Babu, M. M., van der Lee, R., de Groot, N. S. and Gsponer, J. (2011) 'Intrinsically Disordered Proteins: Regulation and Disease.', *Current Opinion in Structural Biology*, 21(3), pp. 432–440. doi: 10.1016/j.sbi.2011.03.011.

Bachmann, A., Wildemann, D., Praetorius, F., Fischer, G. and Kiefhaber, T. (2011) 'Mapping Backbone and Side-chain Interactions in the Transition State of a Coupled Protein Folding and Binding Reaction.', *Proceedings of the National Academy of Sciences*, 108(10), pp. 3952–3957. doi: 10.1073/pnas.1012668108.

Bah, A. and Forman-Kay, J. D. (2016) 'Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications.', *Journal of Biological Chemistry*, 291(13), pp. 6696–6705. doi: 10.1074/jbc.R115.695056.

Bah, A., Vernon, R. M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L. E. and Forman-Kay, J. D. (2014) 'Folding of an Intrinsically Disordered Protein by Phosphorylation as a Regulatory Switch.', *Nature*, 519(7541), pp. 106–109. doi: 10.1038/nature13999.

Balasubramaniam, D. and Komives, E. A. (2013) 'Hydrogen-Exchange Mass Spectrometry for the Study of Intrinsic Disorder in Proteins.', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(6), pp. 1202–1209. doi: 10.1016/j.bbapap.2012.10.009.

Banci, L., Bertini, I., Cefaro, C., Ciofi-Baffoni, S. and Gallo, A. (2011) 'Functional Role of Two Interhelical Disulfide Bonds in Human Cox17 Protein from a Structural Perspective.', *Journal of Biological Chemistry*, 286(39), pp. 34382–34390. doi: 10.1074/jbc.M111.246223.

Banci, L., Bertini, I., Ciofi-Baffoni, S., Hadjiloi, T., Martinelli, M. and Palumaa, P. (2008) 'Mitochondrial Copper(I) Transfer from Cox17 to Sco1 is Coupled to Electron Transfer.', *Proceedings of the National Academy of Sciences*, 105(19), pp. 6803–6808. doi: 10.1073/pnas.0800019105.

Banci, L., Bertini, I., Ciofi-Baffoni, S., Janicka, A., Martinelli, M., Kozlowski, H. and Palumaa, P. (2008) 'A Structural-Dynamical Characterization of Human Cox17.', *Journal of Biological Chemistry*, 283(12), pp. 7912–7920. doi: 10.1074/jbc.M708016200.

Barnum, C. J. and Tansey, M. G. (2012) 'Neuroinflammation and Non-motor Symptoms: The Dark Passenger of Parkinson's Disease?', *Current Neurology and Neuroscience Reports*, 12(4), pp. 350–358. doi: 10.1007/s11910-012-0283-6.

Barrick, D. (2009) 'Biological Regulation via Ankyrin Repeat Folding.', *ACS Chemical Biology*, 4(1), pp. 19–22. doi: 10.1021/cb900003f.

Baxa, M. C., Haddadian, E. J., Jumper, J. M., Freed, K. F. and Sosnick, T. R. (2014) 'Loss of Conformational Entropy in Protein Folding Calculated using Realistic Ensembles and its Implications for NMR-based Calculations.', *Proceedings of the National Academy of Sciences*, 111(43), pp. 15396–15401. doi: 10.1073/pnas.1407768111.

Behrens, J., von Kries, J. P., Kühl, M., Bruhn, L., Wedlich, D., Grosschedl, R. and Birchmeier, W. (1996) 'Functional Interaction of β-catenin with the Transcription Factor LEF-1.', *Nature*, 382(6592), pp. 638–642. doi: 10.1038/382638a0.

Bella, J., Hindle, K. L., McEwan, P. A. and Lovell, S. C. (2008) 'The Leucine-Rich Repeat Structure.', *Cellular and Molecular Life Sciences*, 65(15), pp. 2307–33. doi: 10.1007/s00018-008-8019-0.

Bergqvist, S., Alverdi, V., Mengel, B., Hoffmann, A., Ghosh, G. and Komives, E. A. (2009) 'Kinetic Enhancement of NF-κB-DNA dissociation by IκBα.', *Proceedings of the National Academy of Sciences*, 106(46), pp. 19328–19333. doi: 10.1073/pnas.0908797106.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) 'The Protein Data Bank.', *Nucleic Acids Research*, 28(1), pp. 235–42. doi: 10.1093/nar/28.1.235.

Berwick, D. C. and Harvey, K. (2012) 'LRRK2 Functions as a Wnt Signaling Scaffold, Bridging Cytosolic Proteins and Membrane-Localized LRP6.', *Human Molecular Genetics*, 21(22), pp. 4966–79. doi: 10.1093/hmg/dds342.

Berwick, D. C. and Harvey, K. (2013) 'LRRK2: an Éminence Grise of Wnt-mediated Neurogenesis?', *Frontiers in Cellular Neuroscience*, 7, p. 82. doi: 10.3389/fncel.2013.00082.

Bienz, M. and Clevers, H. (2000) 'Linking Colorectal Cancer to Wnt Signaling.', *Cell*, 103(2), pp. 311–320. doi: 10.1016/S0092-8674(00)00122-7.

Bilic, J., Huang, Y.-L., Davidson, G., Zimmermann, T., Cruciat, C.-M., Bienz, M. and Niehrs, C. (2007) 'Wnt Induces LRP6 Signalosomes and Promotes Dishevelled-dependent LRP6 Phosphorylation.', *Science*, 316(5831), pp. 1619–1622. doi: 10.1126/science.1137065.

Birnboim, H. C. and Doly, J. (1979) 'A Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid DNA.', *Nucleic Acids Research*, 7(6), pp. 1513–23. doi: 10.1093/nar/7.6.1513.

Biskup, S., Moore, D. J., Celsi, F., Higashi, S., West, A. B., Andrabi, S. A., Kurkinen, K., Yu, S.-W., Savitt, J. M., Waldvogel, H. J., Faull, R. L. M., Emson, P. C., Torp, R., Ottersen, O. P., Dawson, T. M. and Dawson, V. L. (2006) 'Localization of LRRK2 to Membranous and Vesicular Structures in Mammalian Brain.', *Annals of Neurology*, 60(5), pp. 557–569. doi: 10.1002/ana.21019.

Blaber, M. and Lee, J. (2012) 'Designing Proteins from Simple Motifs: Opportunities in Top-down Symmetric Deconstruction.', *Current Opinion in Structural Biology*, 22(4), pp. 442–450. doi: 10.1016/j.sbi.2012.05.008.

Blitzer, J. T. and Nusse, R. (2006) 'A Critical Role for Endocytosis in Wnt Signaling.', *BMC Cell Biology*, 7(1), p. 28. doi: 10.1186/1471-2121-7-28.

Bobrovnik, S. A. (2007) 'The Influence of Rigid or Flexible Linkage between Two Ligands on the Effective Affinity and Avidity for Reversible Interactions with Bivalent Receptors.', *Journal of Molecular Recognition*, 20(4), pp. 253–262. doi: 10.1002/jmr.836.

de Boer, H. A., Comstock, L. J. and Vasser, M. (1983) 'The tac Promoter: A Functional Hybrid Derived from the trp and lac Promoters.', *Proceedings of the National Academy of Sciences*, 80(1), pp. 21–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=393301&tool=pmcentrez&rendertype=abstract (Accessed: 1 June 2014).

Bogan, A. A. and Thorn, K. S. (1998) 'Anatomy of Hot Spots in Protein Interfaces.', *Journal of Molecular Biology*, 280(1), pp. 1–9. doi: 10.1006/jmbi.1998.1843.

Bokor, M., Csizmók, V., Kovács, D., Bánki, P., Friedrich, P., Tompa, P. and Tompa, K. (2005) 'NMR Relaxation Studies on the Hydrate Layer of Intrinsically Unstructured Proteins.', *Biophysical Journal*, 88(3), pp. 2030–2037. doi: 10.1529/biophysj.104.051912.

Bordelon, T., Montegudo, S. K., Pakhomova, S., Oldham, M. L. and Newcomer, M. E. (2004) 'A Disorder to Order Transition Accompanies Catalysis in Retinaldehyde Dehydrogenase Type II.', *Journal of Biological Chemistry*, 279(41), pp. 43085–43091. doi: 10.1074/jbc.M406139200.

Borg, M., Mittag, T., Pawson, T., Tyers, M., Forman-Kay, J. D. and Chan, H. S. (2007) 'Polyelectrostatic Interactions of Disordered Ligands Suggest a Physical Basis for Ultrasensitivity.', *Proceedings of the National Academy of Sciences*, 104(23), pp. 9650–5. doi: 10.1073/pnas.0702580104.

Bradley, C. M. and Barrick, D. (2006) 'The Notch Ankyrin Domain Folds via a Discrete, Centralized Pathway.', *Structure*, 14(8), pp. 1303–1312. doi: 10.1016/j.str.2006.06.013.

Brantjes, H., Roose, J., van De Wetering, M. and Clevers, H. (2001) 'All Tcf HMG Box Transcription Factors Interact with Groucho-related Co-repressors.', *Nucleic Acids Research*, 29(7), pp. 1410–9. doi: 10.1093/nar/29.7.1410.

Breeden, L. and Nasmyth, K. (1987) 'Similarity Between Cell-Cycle Genes of Budding Yeast and Fission Yeast and the Notch Gene of Drosophila.', *Nature*, 329(6140), pp. 651–654. doi: 10.1038/329651a0.

Breslauer, K. J., Frank, R., Blöcker, H. and Marky, L. A. (1986) 'Predicting DNA duplex stability from the base sequence.', *Proceedings of the National Academy of Sciences of the United States of America*, 83(11), pp. 3746–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3459152.

Brundin, P. and Melki, R. (2017) 'Prying into the Prion Hypothesis for Parkinson's Disease.', *The Journal of Neuroscience*, 37(41), pp. 9808–9818. doi: 10.1523/JNEUROSCI.1788-16.2017.

Brundin, P., Melki, R. and Kopito, R. (2010) 'Prion-like Transmission of Protein Aggregates in Neurodegenerative Diseases.', *Nature Reviews Molecular Cell Biology*, 11(4), pp. 301–307. doi: 10.1038/nrm2873.

Burling, F. T., Weis, W. I., Flaherty, K. M. and Br nger, A. T. (1996) 'Direct Observation of Protein Solvation and Discrete Disorder with Experimental Crystallographic Phases.', *Science*, 271(5245), pp. 72–77. doi: 10.1126/science.271.5245.72.

Burré, J. (2015) 'The Synaptic Function of α-Synuclein.', *Journal of Parkinson's Disease*, 5(4), pp. 699–713. doi: 10.3233/JPD-150642.

Caballol, N., Martí, M. J. and Tolosa, E. (2007) 'Cognitive Dysfunction and Dementia in Parkinson Disease.', *Movement Disorders: Official Journal of the Movement Disorder Society*, 22 Suppl 1, pp. S358-66. doi: 10.1002/mds.21677.

Cadigan, K. M. and Waterman, M. L. (2012) 'TCF/LEFs and Wnt Signaling in the Nucleus.', *Cold Spring Harbor Perspectives in Biology*, 4(11), pp. a007906–a007906. doi: 10.1101/cshperspect.a007906.

Carlson, M. A., Haddad, B. G., Weis, A. J., Blackwood, C. S., Shelton, C. D., Wuerth, M. E., Walter, J. D. and Spiegel, P. C. (2017) 'Ribosomal Protein L7/L12 is Required for GTPase Translation Factors EF-G, RF3, and IF2 to Bind in their GTP State to 70S Ribosomes.', *The FEBS Journal*, 284(11), pp. 1631–1643. doi: 10.1111/febs.14067.

Castelo-Branco, G., Andersson, E. R., Minina, E., Sousa, K. M., Ribeiro, D., Kokubu, C., Imai, K., Prakash, N., Wurst, W. and Arenas, E. (2009) 'Delayed Dopaminergic Neuron Differentiation in LRP6

Mutant Mice.', *Developmental Dynamics*, 239(1), p. NA-NA. doi: 10.1002/dvdy.22094.

Castelo-Branco, G., Wagner, J., Rodriguez, F. J., Kele, J., Sousa, K., Rawal, N., Pasolli, H. A., Fuchs, E., Kitajewski, J. and Arenas, E. (2003) 'Differential Regulation of Midbrain Dopaminergic Neuron Development by Wnt-1, Wnt-3a, and Wnt-5a.', *Proceedings of the National Academy of Sciences*, 100(22), pp. 12747–12752. doi: 10.1073/pnas.1534900100.

Chemes, L. B., Sánchez, I. E. and de Prat-Gay, G. (2011) 'Kinetic Recognition of the Retinoblastoma Tumor Suppressor by a Specific Protein Target.', *Journal of Molecular Biology*, 412(2), pp. 267–284. doi: 10.1016/j.jmb.2011.07.015.

Chen, G., Fernandez, J., Mische, S. and Courey, A. J. (1999) 'A Functional Interaction between the Histone Deacetylase Rpd3 and the Corepressor Groucho in Drosophila Development.', *Genes & Development*, 13(17), pp. 2218–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10485845.

Chen, M., Philipp, M., Wang, J., Premont, R. T., Garrison, T. R., Caron, M. G., Lefkowitz, R. J. and Chen, W. (2009) 'G Protein-coupled Receptor Kinases Phosphorylate LRP6 in the Wnt Pathway.', *Journal of Biological Chemistry*, 284(50), pp. 35040–35048. doi: 10.1074/jbc.M109.047456.

Choi, H.-J., Gross, J. C., Pokutta, S. and Weis, W. I. (2009) 'Interactions of Plakoglobin and β-catenin with Desmosomal Cadherins.', *Journal of Biological Chemistry*, 284(46), pp. 31776–31788. doi: 10.1074/jbc.M109.047928.

Choi, H.-J., Huber, A. H. and Weis, W. I. (2006) 'Thermodynamics of β-catenin-ligand Interactions: The Roles of the N- and C-terminal Tails in Modulating Binding Affinity.', *Journal of Biological Chemistry*, 281(2), pp. 1027–1038. doi: 10.1074/jbc.M511338200.

Chook, Y. M. and Süel, K. E. (2011) 'Nuclear Import by Karyopherin-βs: Recognition and Inhibition.', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1813(9), pp. 1593–1606. doi: 10.1016/j.bbamcr.2010.10.014.

Choy, W.-Y. and Forman-Kay, J. D. (2001) 'Calculation of Ensembles of Structures Representing the Unfolded State of an SH3 Domain.', *Journal of Molecular Biology*, 308(5), pp. 1011–1032. doi: 10.1006/jmbi.2001.4750.

Clackson, T. and Wells, J. (1995) 'A Hot Spot of Binding Energy in a Hormone-receptor Interface.', *Science*, 267(5196), pp. 383–386. doi: 10.1126/science.7529940.

Clevers, H. and Nusse, R. (2012) 'Wnt/β-catenin Signaling and Disease.', *Cell*, 149(6), pp. 1192–1205. doi: 10.1016/j.cell.2012.05.012.

Coates, J. (2003) 'Armadillo Repeat Proteins: Beyond the Animal Kingdom.', *Trends in Cell Biology*, 13(9), pp. 463–471. doi: 10.1016/S0962-8924(03)00167-3.

Cobine, P. A., Pierrel, F. and Winge, D. R. (2006) 'Copper Trafficking to the Mitochondrion and Assembly of Copper Metalloenzymes.', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1763(7), pp. 759–772. doi: 10.1016/j.bbamcr.2006.03.002.

Cockman, M. E., Webb, J. D., Kramer, H. B., Kessler, B. M. and Ratcliffe, P. J. (2009) 'Proteomics-based Identification of Novel Factor Inhibiting Hypoxia-inducible Factor (FIH) Substrates Indicates Widespread Asparaginyl Hydroxylation of Ankyrin Repeat Domain-containing Proteins.', *Molecular & Cellular Proteomics*, 8(3), pp. 535–546. doi: 10.1074/mcp.M800340-MCP200.

Coleman, M. L., McDonough, M. A., Hewitson, K. S., Coles, C., Mecinović, J., Edelmann, M., Cook, K. M., Cockman, M. E., Lancaster, D. E., Kessler, B. M., Oldham, N. J., Ratcliffe, P. J. and Schofield, C. J. (2007) 'Asparaginyl Hydroxylation of the Notch Ankyrin Repeat Domain by Factor Inhibiting Hypoxia-inducible Factor.', *Journal of Biological Chemistry*, 282(33), pp. 24027–24038. doi: 10.1074/jbc.M704102200.

Conacci-Sorrell, M., Zhurinsky, J. and Ben-Ze'ev, A. (2002) 'The Cadherin-catenin Adhesion System in Signaling and Cancer.', *Journal of Clinical Investigation*, 109(8), pp. 987–991. doi: 10.1172/JCI15429.

Conway, K. A., Lee, S.-J., Rochet, J.-C., Ding, T. T., Williamson, R. E. and Lansbury, P. T. (2000) 'Acceleration of Oligomerization, not Fibrillization, is a Shared Property of Both α-synuclein Mutations Linked to Early-onset Parkinson's Disease: Implications for Pathogenesis and Therapy.', *Proceedings of the National Academy of Sciences*, 97(2), pp. 571–576. doi: 10.1073/pnas.97.2.571.

Cookson, M. R. (2010) 'The Role of Leucine-Rich Repeat Kinase 2 (LRRK2) in Parkinson's Disease.', *Nature reviews. Neuroscience*, 11(12), pp. 791–797. doi: 10.1038/nrn2935.

Courtemanche, N. and Barrick, D. (2008) 'The Leucine-Rich Repeat Domain of Internalin B Folds along a Polarized N-Terminal Pathway.', *Structure*, 16(5), pp. 705–714. doi: 10.1016/j.str.2008.02.015.

Crawford, H. C., Fingleton, B. M., Rudolph-Owen, L. A., Goss, K. J. H., Rubinfeld, B., Polakis, P. and Matrisian, L. M. (1999) 'The Metalloproteinase Matrilysin is a Target of β-catenin Transactivation in Intestinal Tumors.', *Oncogene*, 18(18), pp. 2883–2891. doi: 10.1038/sj.onc.1202627.

Crothers, D. M. and Metzger, H. (1972) 'The Influence of Polyvalency on the Binding Properties of Antibodies.', *Immunochemistry*, 9(3), pp. 341–57. doi: 10.1016/0019-2791(72)90097-3.

Cruciat, C.-M., Ohkawara, B., Acebron, S. P., Karaulanov, E., Reinhard, C., Ingelfinger, D., Boutros, M. and Niehrs, C. (2010) 'Requirement of Prorenin Receptor and Vacuolar H+-ATPase-mediated Acidification for Wnt Signaling.', *Science*, 327(5964), pp. 459–463. doi: 10.1126/science.1179802.

Csizmók, V., Bokor, M., Bánki, P., Klement, É., Medzihradszky, K. F., Friedrich, P., Tompa, K. and Tompa, P. (2005) 'Primary Contact Sites in Intrinsically Unstructured Proteins: The Case of Calpastatin and Microtubule-associated Protein 2.', *Biochemistry*, 44(10), pp. 3955–3964. doi: 10.1021/bi047817f.

Da Cunha, C., Silva, M. H. C., Wietzikoski, S., Wietzikoski, E. C., Ferro, M. M., Kouzmine, I. and Canteras, N. S. (2006) 'Place Learning Strategy of Substantia Nigra Pars Compacta-lesioned Rats.', *Behavioral Neuroscience*, 120(6), pp. 1279–84. doi: 10.1037/0735-7044.120.6.1279.

Cuniasse, P., Tavares, P., Orlova, E. V and Zinn-Justin, S. (2017) 'Structures of Biomolecular Complexes by Combination of NMR and cryoEM Methods.', *Current Opinion in Structural Biology*, 43, pp. 104–113. doi: 10.1016/j.sbi.2016.12.008.

Daniels, D. L., Eklof Spink, K. and Weis, W. I. (2001) 'β-catenin: Molecular Plasticity and Drug Design.', *Trends in Biochemical Sciences*, 26(11), pp. 672–678. doi: 10.1016/S0968-0004(01)01952-1.

Das, R. K. and Pappu, R. V (2013) 'Conformations of Intrinsically Disordered Proteins are Influenced by Linear Sequence Distributions of Oppositely Charged Residues.', *Proceedings of the National Academy of Sciences*, 110(33), pp. 13392–7. doi: 10.1073/pnas.1304749110.

Dauer, W. and Przedborski, S. (2003) 'Parkinson's Disease: Mechanisms and Models.', *Neuron*, 39(6), pp. 889–909. doi: 10.1016/S0896-6273(03)00568-3.

Davey, N. E., Cyert, M. S. and Moses, A. M. (2015) 'Short Linear Motifs - ex nihilo Evolution of Protein Regulation.', *Cell Communication and Signaling*, 13(1), p. 43. doi: 10.1186/s12964-015-0120-z.

Davidson, G. and Niehrs, C. (2010) 'Emerging Links between CDK Cell Cycle Regulators and Wnt Signaling.', *Trends in Cell Biology*, 20(8), pp. 453–460. doi: 10.1016/j.tcb.2010.05.002.

Davidson, G., Shen, J., Huang, Y.-L., Su, Y., Karaulanov, E., Bartscherer, K., Hassler, C., Stannek, P., Boutros, M. and Niehrs, C. (2009) 'Cell Cycle Control of Wnt Receptor Activation.', *Developmental Cell*, 17(6), pp. 788–799. doi: 10.1016/j.devcel.2009.11.006.

Davidson, G., Wu, W., Shen, J., Bilic, J., Fenger, U., Stannek, P., Glinka, A. and Niehrs, C. (2005) 'Casein Kinase 1γ Couples Wnt Receptor Activation to Cytoplasmic Signal Transduction.', *Nature*, 438(7069), pp. 867–872. doi: 10.1038/nature04170.

Davie, C. A. (2008) 'A Review of Parkinson's Disease.', *British Medical Bulletin*, 86, pp. 109–27. doi: 10.1093/bmb/ldn013.

Dawson, P., Muir, T., Clark-Lewis, I. and Kent, S. (1994) 'Synthesis of Proteins by Native Chemical Ligation.', *Science*, 266(5186), pp. 776–779. doi: 10.1126/science.7973629.

Dedmon, M. M., Patel, C. N., Young, G. B. and Pielak, G. J. (2002) 'FlgM gains Structure in Living Cells.', *Proceedings of the National Academy of Sciences*, 99(20), pp. 12681–12684. doi: 10.1073/pnas.202331299.

Deng, J., Lewis, P. A., Greggio, E., Sluch, E., Beilina, A. and Cookson, M. R. (2008) 'Structure of the ROC Domain from the Parkinson's Disease-associated Leucine-rich Repeat Kinase 2 Reveals a Dimeric GTPase.', *Proceedings of the National Academy of Sciences*, 105(5), pp. 1499–1504.

Devries, I. L., Hampton-Smith, R. J., Mulvihill, M. M., Alverdi, V., Peet, D. J. and Komives, E. A. (2010) 'Consequences of IκBα Hydroxylation by the Factor Inhibiting HIF (FIH).', *FEBS Letters*, 584(23), pp. 4725–4730. doi: 10.1016/j.febslet.2010.10.060.

Dogan, J., Gianni, S. and Jemth, P. (2014) 'The Binding Mechanisms of Intrinsically Disordered Proteins.', *Physical Chemistry Chemical Physics*, 16(14), pp. 6323–31. doi: 10.1039/c3cp54226b.

Dogan, J., Schmidt, T., Mu, X., Engström, Å. and Jemth, P. (2012) 'Fast Association and Slow Transitions in the Interaction between Two Intrinsically Disordered Protein Domains.', *Journal of Biological Chemistry*, 287(41), pp. 34316–34324. doi: 10.1074/jbc.M112.399436.

Dougherty, M. K. and Morrison, D. K. (2004) 'Unlocking the Code of 14-3-3.', *Journal of Cell Science*, 117(Pt 10), pp. 1875–84. doi: 10.1242/jcs.01171.

Drolet, R. E., Sanders, J. M. and Kern, J. T. (2011) 'Leucine-Rich Repeat Kinase 2 (LRRK2) Cellular Biology: A Review of Recent Advances in Identifying Physiological Substrates and Cellular Functions.', *Journal of Neurogenetics*, 25(4), pp. 140–151. doi: 10.3109/01677063.2011.627072.

Drubin, D. G. and Kirschner, M. W. (1986) 'Tau Protein Function in Living Cells.', *The Journal of Cell Biology*, 103(6 Pt 2), pp. 2739–46. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2114585&tool=pmcentrez&rendertype=abstract (Accessed: 24 June 2014).

Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. and Uversky, V. N. (2005) 'Flexible Nets: The Roles of Intrinsic Disorder in Protein Interaction Networks.', *The FEBS Journal*, 272(20), pp. 5129–48. doi: 10.1111/j.1742-4658.2005.04948.x.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. and Obradovic, Z. (2001) 'Intrinsically Disordered Protein.', *Journal of Molecular Graphics & Modelling*, 19(1), pp. 26–59. doi: 10.1016/S1093-3263(00)00138-8.

Dunker, A. K. and Uversky, V. N. (2010) 'Drugs for "Protein Clouds": Targeting Intrinsically Disordered Transcription Factors.', *Current Opinion in Pharmacology*, 10(6), pp. 782–788. doi: 10.1016/j.coph.2010.09.005.

Dyson, H. J. and Wright, P. E. (2002) 'Coupling of Folding and Binding for Unstructured Proteins.', *Current Opinion in Structural Biology*, 12(1), pp. 54–60. doi: 10.1016/S0959-440X(02)00289-0.

Dyson, H. J. and Wright, P. E. (2005) 'Intrinsically Unstructured Proteins and their Functions.', *Nature Reviews Molecular Cell Biology*, 6(3), pp. 197–208. doi: 10.1038/nrm1589.

Dzamko, N., Deak, M., Hentati, F., Reith, A. D., Prescott, A. R., Alessi, D. R. and Nichols, R. J. (2010) 'Inhibition of LRRK2 Kinase Activity Leads to Dephosphorylation of Ser(910)/Ser(935), Disruption of 14-3-3 Binding and Altered Cytoplasmic Localization.', *The Biochemical Journal*, 430(3), pp. 405–13. doi: 10.1042/BJ20100784.

Ebbes, M., Bleymüller, W. M., Cernescu, M., Nölker, R., Brutschy, B. and Niemann, H. H. (2011) 'Fold and Function of the InlB B-repeat.', *Journal of Biological Chemistry*, 286(17), pp. 15496–15506. doi: 10.1074/jbc.M110.189951.

Edelhoch, H. (1967) 'Spectroscopic Determination of Tryptophan and Tyrosine in Proteins.', *Biochemistry*, 6(7), pp. 1948–1954. doi: 10.1021/bi00859a010.

Eklof Spink, K., Fridman, S. G. and Weis, W. I. (2001) 'Molecular Mechanisms of β-catenin Recognition by Adenomatous Polyposis Coli Revealed by the Structure of an APC-β-catenin Complex.', *The EMBO journal*, 20(22), pp. 6203–12. doi: 10.1093/emboj/20.22.6203.

English, L. R., Tilton, E. C., Ricard, B. J. and Whitten, S. T. (2016) 'Intrinsic α Helix Propensities Compact Hydrodynamic Radii in Intrinsically Disordered Proteins.', *Proteins*. doi: 10.1002/prot.25222.

Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T. and Matsushima, N. (2003) 'Structural Principles of Leucine-Rich Repeat (LRR) Proteins.', *Proteins: Structure, Function, and Bioinformatics*, 54(3), pp. 394–403. doi: 10.1002/prot.10605.

van Es, J. H., Haegebarth, A., Kujala, P., Itzkovitz, S., Koo, B.-K., Boj, S. F., Korving, J., van den Born, M., van Oudenaarden, A., Robine, S. and Clevers, H. (2012) 'A Critical Role for the Wnt Effector Tcf4 in Adult Intestinal Homeostatic Self-renewal.', *Molecular and Cellular Biology*, 32(10), pp. 1918–1927. doi: 10.1128/MCB.06288-11.

Fasolini, M., Wu, X., Flocco, M., Trosset, J.-Y., Oppermann, U. and Knapp, S. (2003) 'Hot Spots in Tcf4 for the Interaction with β-catenin.', *Journal of Biological Chemistry*, 278(23), pp. 21092–21098. doi: 10.1074/jbc.M301781200.

Ferreiro, D. U., Cho, S. S., Komives, E. A. and Wolynes, P. G. (2005) 'The Energy Landscape of Modular Repeat Proteins: Topology Determines Folding Mechanism in the Ankyrin Family.', *Journal of Molecular Biology*, 354(3), pp. 679–692. doi: 10.1016/j.jmb.2005.09.078.

Ferreiro, D. U. and Komives, E. A. (2007) 'The plastic landscape of repeat proteins.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), pp. 7735–6. doi: 10.1073/pnas.0702682104.

Fersht, A. R. (1997) 'Nucleation Mechanisms in Protein Folding.', *Current Opinion in Structural Biology*, 7(1), pp. 3–9. doi: 10.1016/S0959-440X(97)80002-4.

Fersht, A. R., Matouschek, A. and Serrano, L. (1992) 'The Folding of an Enzyme. I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding.', *Journal of Molecular Biology*, 224(3), pp. 771–82. doi: 10.1016/0022-2836(92)90561-W.

Fersht, A. R. and Sato, S. (2004) 'Phi-value Analysis and the Nature of Protein-folding Transition States.', *Proceedings of the National Academy of Sciences*, 101(21), pp. 7976–81. doi: 10.1073/pnas.0402684101.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. and Punta, M. (2014) 'Pfam: The Protein Families Database.', *Nucleic Acids Research*. doi: 10.1093/nar/gkt1223.

Fischer, E. (1894) 'Einfluss der Configuration auf die Wirkung der Enzyme', *Berichte der deutschen chemischen Gesellschaft*, 27(3), pp. 2985–2993. doi: 10.1002/cber.18940270364.

Forrer, P., Binz, H. K., Stumpp, M. T. and Plückthun, A. (2004) 'Consensus Design of Repeat Proteins.', *ChemBioChem*, 5(2), pp. 183–189. doi: 10.1002/cbic.200300762.

Fürstner, A. and Langemann, K. (1996) 'A Concise Total Synthesis of Dactylol via Ring Closing Metathesis.', *The Journal of Organic Chemistry*, 61(25), pp. 8746–8749. doi: 10.1021/jo961600c.

Fuxreiter, M., Tompa, P. and Simon, I. (2007) 'Local Structural Disorder Imparts Plasticity on Linear Motifs.', *Bioinformatics*, 23(8), pp. 950–956. doi: 10.1093/bioinformatics/btm035.

Gaboriau, D. C. A., Rowling, P. J. E., Morrison, C. G. and Itzhaki, L. S. (2015) 'Protein Stability versus Function: Effects of Destabilizing Missense Mutations on BRCA1 DNA Repair Activity.', *Biochemical Journal*, 466(3), pp. 613–624. doi: 10.1042/BJ20141077.

Gail, R., Frank, R. and Wittinghofer, A. (2005) 'Systematic Peptide Array-based Delineation of the Differential β-catenin Interaction with Tcf4, E-cadherin, and Adenomatous Polyposis Coli.', *The Journal of Biological Chemistry*, 280(8), pp. 7107–17. doi: 10.1074/jbc.M410215200.

Gamini, R., Han, W., Stone, J. E. and Schulten, K. (2014) 'Assembly of Nsp1 Nucleoporins Provides Insight into Nuclear Pore Complex Gating.', *PLoS Computational Biology*. Edited by H. Grubmüller, 10(3), p. e1003488. doi: 10.1371/journal.pcbi.1003488.

Gandhi, P. N., Wang, X., Zhu, X., Chen, S. G. and Wilson-Delfosse, A. L. (2008) 'The ROC Domain of Leucine-rich Repeat Kinase 2 is Sufficient for Interaction with Microtubules.', *Journal of Neuroscience Research*, 86(8), pp. 1711–20. doi: 10.1002/jnr.21622.

Gao, C. and Chen, Y.-G. (2010) 'Dishevelled: The Hub of Wnt Signaling.', *Cellular Signalling*, 22(5), pp. 717–27. doi: 10.1016/j.cellsig.2009.11.021.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) *The Proteomics Protocols Handbook - Chapter 52: Protein Identification and Analysis Tools on the ExPASy Server*, *The Proteomics Protocols Handbook*. Edited by J. M. Walker. Totowa, NJ: Humana Press. doi: 10.1385/1592598900.

Gibbs, E. B. and Showalter, S. A. (2015) 'Quantitative Biophysical Characterization of Intrinsically Disordered Proteins.', *Biochemistry*, p. 150206124301003. doi: 10.1021/bi501460a.

Gillardon, F. (2009) 'Leucine-rich Repeat Kinase 2 Phosphorylates Brain Tubulin-β Isoforms and Modulates Microtubule Stability - a Point of Convergence in Parkinsonian Neurodegeneration?', *Journal of Neurochemistry*, 110(5), pp. 1514–22. doi: 10.1111/j.1471-4159.2009.06235.x.

Gloeckner, C. J., Schumacher, A., Boldt, K. and Ueffing, M. (2009) 'The Parkinson Disease-associated Protein Kinase LRRK2 Exhibits MAPKKK Activity and Phosphorylates MKK3/6 and MKK4/7, in vitro.', *Journal of Neurochemistry*, 109(4), pp. 959–968. doi: 10.1111/j.1471-4159.2009.06024.x.

Gógl, G., Alexa, A., Kiss, B., Katona, G., Kovács, M., Bodor, A., Reményi, A. and Nyitray, L. (2016) 'Structural Basis of Ribosomal S6 Kinase 1 (RSK1) Inhibition by S100B Protein.', *Journal of Biological*

Chemistry, 291(1), pp. 11–27. doi: 10.1074/jbc.M115.684928.

Goldberg, J. M. and Baldwin, R. L. (1998) 'Kinetic Mechanism of a Partial Folding Reaction. 1. Properties of the Reaction and Effects of Denaturants.', *Biochemistry*, 37(8), pp. 2546–2555. doi: 10.1021/bi972402y.

Gontero, B. and Maberly, S. C. (2012) 'An Intrinsically Disordered Protein, CP12: Jack of all Trades and Master of the Calvin Cycle', *Biochemical Society Transactions*, 40(5), pp. 995–999. doi: 10.1042/BST20120097.

Gordon, M. D. and Nusse, R. (2006) 'Wnt Signaling: Multiple Pathways, Multiple Receptors, and Multiple Transcription Factors.', *Journal of Biological Chemistry*, 281(32), pp. 22429–22433. doi: 10.1074/jbc.R600015200.

Görlich, D., Panté, N., Kutay, U., Aebi, U. and Bischoff, F. R. (1996) 'Identification of Different Roles for RanGDP and RanGTP in Nuclear Protein Import.', *The EMBO Journal*, 15(20), pp. 5584–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8896452.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. and Lopez, R. (2010) 'A New Bioinformatics Analysis Tools Framework at EMBL-EBI.', *Nucleic Acids Research*, 38(Web Server issue), pp. W695-9. doi: 10.1093/nar/gkq313.

Graham, T. A., Clements, W. K., Kimelman, D. and Xu, W. (2002) 'The Crystal Structure of the β-catenin/ICAT Complex Reveals the Inhibitory Mechanism of ICAT.', *Molecular Cell*, 10(3), pp. 563–71. doi: 10.1016/S1097-2765(02)00637-8.

Graham, T. A., Ferkey, D. M., Mao, F., Kimelman, D. and Xu, W. (2001) 'Tcf4 can Specifically Recognize β-catenin using Alternative Conformations.', *Nature Structural Biology*, 8(12), pp. 1048–52. doi: 10.1038/nsb718.

Graham, T. A., Weaver, C., Mao, F., Kimelman, D. and Xu, W. (2000) 'Crystal Structure of a β-catenin/Tcf Complex.', *Cell*, 103(6), pp. 885–96. doi: 10.1016/S0092-8674(00)00192-6.

Grossmann, T. N., Yeh, J. T.-H., Bowman, B. R., Chu, Q., Moellering, R. E. and Verdine, G. L. (2012) 'Inhibition of Oncogenic Wnt Signaling Through Direct Targeting of β-catenin.', *Proceedings of the National Academy of Sciences*, 109(44), pp. 17942–7. doi: 10.1073/pnas.1208396109.

Grove, T., Cortajarena, A. and Regan, L. (2008) 'Ligand Binding by Repeat Proteins: Natural and Designed.', *Current Opinion in Structural Biology*, 18(4), pp. 507–515. doi: 10.1016/j.sbi.2008.05.008.

Grove, T. Z., Forster, J., Pimienta, G., Dufresne, E. and Regan, L. (2012) 'A Modular Approach to the Design of Protein-based Smart Gels.', *Biopolymers*, 97(7), pp. 508–517. doi: 10.1002/bip.22033.

Grove, T. Z., Regan, L. and Cortajarena, A. L. (2013) 'Nanostructured Functional Films from Engineered Repeat Proteins.', *Journal of The Royal Society Interface*, 10(83), pp. 20130051–20130051. doi: 10.1098/rsif.2013.0051.

Grucza, R. A., Bradshaw, J. M., Mitaxov, V. and Waksman, G. (2000) 'Role of Electrostatic Interactions in SH2 Domain Recognition: Salt-dependence of Tyrosyl-phosphorylated Peptide Binding to the Tandem SH2 Domain of the Syk Kinase and the Single SH2 Domain of the Src Kinase.', *Biochemistry*, 39(33), pp. 10072–10081. doi: 10.1021/bi000891n.

Guder, C., Philipp, I., Lengfeld, T., Watanabe, H., Hobmayer, B. and Holstein, T. W. (2006) 'The Wnt Code: Cnidarians Signal the Way.', *Oncogene*, 25(57), pp. 7450–7460. doi: 10.1038/sj.onc.1210052.

Guttenplan, A. G. M., Young, L. J., Matak-Vinkovic, D., Kaminski, C. F., Knowles, T. P. J. and Itzhaki, L. S. (2017) 'Amyloid Fibrils as Nanoscale Click-Reactive Scaffolds.', *Currently under review*.

Ha, N.-C., Tonozuka, T., Stamos, J. L., Choi, H.-J. and Weis, W. I. (2004) 'Mechanism of Phosphorylation-dependent Binding of APC to β-catenin and its Role in β-catenin Degradation.', *Molecular Cell*, 15(4), pp. 511–21. doi: 10.1016/j.molcel.2004.08.010.

Hamaguchi, K. and Geiduschek, E. P. (1962) 'The Effect of Electrolytes on the Stability of the Deoxyribonucleate Helix.', *Journal of the American Chemical Society*, 84(8), pp. 1329–1338. doi: 10.1021/ja00867a001.

Hammoudeh, D. I., Follis, A. V., Prochownik, E. V and Metallo, S. J. (2009) 'Multiple Independent Binding Sites for Small-molecule Inhibitors on the Oncoprotein c-Myc.', *Journal of the American Chemical Society*, 131(21), pp. 7390–7401. doi: 10.1021/ja900616b.

Harris, T. J. C. and Tepass, U. (2010) 'Adherens Junctions: from Molecules to Morphogenesis.', *Nature Reviews Molecular Cell Biology*, 11(7), pp. 502–514. doi: 10.1038/nrm2927.

Hatzis, P., van der Flier, L. G., van Driel, M. A., Guryev, V., Nielsen, F., Denissov, S., Nijman, I. J., Koster, J., Santo, E. E., Welboren, W., Versteeg, R., Cuppen, E., van de Wetering, M., Clevers, H. and Stunnenberg, H. G. (2008) 'Genome-wide Pattern of TCF7L2/TCF4 Chromatin Occupancy in Colorectal Cancer Cells.', *Molecular and Cellular Biology*, 28(8), pp. 2732–2744. doi: 10.1128/MCB.02175-07.

Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M. and

Iakoucheva, L. M. (2006) 'Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes.', *PLoS Computational Biology*, 2(8), p. e100. doi: 10.1371/journal.pcbi.0020100.

Hazra, A., Fuchs, C. S., Chan, A. T., Giovannucci, E. L. and Hunter, D. J. (2008) 'Association of the TCF7L2 Polymorphism with Colorectal Cancer and Adenoma Risk.', *Cancer Causes & Control*, 19(9), pp. 975–980. doi: 10.1007/s10552-008-9164-3.

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N. and Dunker, A. K. (2009) 'Predicting Intrinsic Disorder in Proteins: an Overview.', *Cell Research*, 19(8), pp. 929–949. doi: 10.1038/cr.2009.87.

Heger, A. and Holm, L. (2000) 'Rapid Automatic Detection and Alignment of Repeats in Protein Sequences.', *Proteins*, 41(2), pp. 224–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10966575 (Accessed: 18 June 2014).

Hemsath, L., Dvorsky, R., Fiegen, D., Carlier, M.-F. and Ahmadian, M. R. (2005) 'An Electrostatic Steering Mechanism of Cdc42 Recognition by Wiskott-Aldrich Syndrome Proteins.', *Molecular Cell*, 20(2), pp. 313–324. doi: 10.1016/j.molcel.2005.08.036.

Heringa, J. (1998) 'Detection of Internal Repeats: How Common are they?', *Current Opinion in Structural Biology*, 8(3), pp. 338–345. doi: 10.1016/S0959-440X(98)80068-7.

Herzig, M. C., Kolly, C., Persohn, E., Theil, D., Schweizer, T., Hafner, T., Stemmelen, C., Troxler, T. J., Schmid, P., Danner, S., Schnell, C. R., Mueller, M., Kinzel, B., Grevot, A., Bolognani, F., Stirn, M., Kuhn, R. R., Kaupmann, K., van der Putten, P. H., Rovelli, G. and Shimshek, D. R. (2011) 'LRRK2 Protein Levels are Determined by Kinase Function and are Crucial for Kidney and Lung Homeostasis in Mice.', *Human Molecular Genetics*, 20(21), pp. 4209–23. doi: 10.1093/hmg/ddr348.

Herzig, M., Savarese, F., Novatchkova, M., Semb, H. and Christofori, G. (2007) 'Tumor Progression Induced by the Loss of E-cadherin Independent of β-catenin/Tcf-mediated Wnt Signaling.', *Oncogene*, 26(16), pp. 2290–2298. doi: 10.1038/sj.onc.1210029.

Himo, F., Lovell, T., Hilgraf, R., Rostovtsev, V. V, Noodleman, L., Sharpless, K. B. and Fokin, V. V (2005) 'Copper(I)-catalyzed Synthesis of Azoles. DFT Study Predicts Unprecedented Reactivity and Intermediates.', *Journal of the American Chemical Society*, 127(1), pp. 210–216. doi: 10.1021/ja0471525.

Hodge, G. K. and Butcher, L. L. (1980) 'Pars Compacta of the Substantia Nigra Modulates Motor Activity but is not Involved Importantly in Regulating Food and Water Intake..', *Naunyn-*

Schmiedeberg's Archives of Pharmacology, 313(1), pp. 51–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7207636 (Accessed: 22 June 2014).

Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. and Pappu, R. V (2017) 'CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins.', *Biophysical Journal*, 112(1), pp. 16–21. doi: 10.1016/j.bpj.2016.11.3200.

Horng, Y.-C., Cobine, P. A., Maxfield, A. B., Carr, H. S. and Winge, D. R. (2004) 'Specific Copper Transfer from the Cox17 Metallochaperone to Both Sco1 and Cox11 in the Assembly of Yeast Cytochrome c Oxidase.', *Journal of Biological Chemistry*, 279(34), pp. 35334–35340. doi: 10.1074/jbc.M404747200.

Hough, L. E., Dutta, K., Sparks, S., Temel, D. B., Kamal, A., Tetenbaum-Novatt, J., Rout, M. P. and Cowburn, D. (2015) 'The Molecular Mechanism of Nuclear Transport Revealed by Atomic-scale Measurements.', *eLife*, 4. doi: 10.7554/eLife.10027.

Hrckulak, D., Kolar, M., Strnad, H. and Korinek, V. (2016) 'TCF/LEF Transcription Factors: An Update from the Internet Resources.', *Cancers*, 8(7), p. 70. doi: 10.3390/cancers8070070.

Hsu, C. H., Chan, D., Greggio, E., Saha, S., Guillily, M. D., Ferree, A., Raghavan, K., Shen, G. C., Segal, L., Ryu, H., Cookson, M. R. and Wolozin, B. (2010) 'MKK6 Binds and Regulates Expression of Parkinson's Disease-related Protein LRRK2.', *Journal of Neurochemistry*, 112(6), pp. 1593–1604. doi: 10.1111/j.1471-4159.2010.06568.x.

Hsu, C. H., Chan, D. and Wolozin, B. (2010) 'LRRK2 and the Stress Response: Interaction with MKKs and JNK-interacting Proteins.', *Neurodegenerative Diseases*, 7(1–3), pp. 68–75. doi: 10.1159/000285509.

Huang, A. and Stultz, C. M. (2008) 'The Effect of a ΔK280 Mutation on the Unfolded State of a Microtubule-binding Repeat in Tau.', *PLoS Computational Biology*. Edited by K. Dunker, 4(8), p. e1000155. doi: 10.1371/journal.pcbi.1000155.

Huber, A. H., Nelson, W. J. and Weis, W. I. (1997) 'Three-dimensional Structure of the Armadillo Repeat Region of β-catenin.', *Cell*, 90(5), pp. 871–882. doi: 10.1016/S0092-8674(00)80352-9.

Huber, A. H. and Weis, W. I. (2001) 'The Structure of the β-catenin/E-cadherin Complex and the Molecular Basis of Diverse Ligand Recognition by β-catenin.', *Cell*, 105(3), pp. 391–402. doi: 10.1016/S0092-8674(01)00330-0.

Hutton, R. D., Wilkinson, J., Faccin, M., Sivertsson, E. M., Pelizzola, A., Lowe, A. R., Bruscolini, P. and

Itzhaki, L. S. (2015) 'Mapping the Topography of a Protein Energy Landscape.', *Journal of the American Chemical Society*, 137(46), pp. 14610–14625. doi: 10.1021/jacs.5b07370.

Hwang, C. J., Choi, D.-Y., Jung, Y. Y., Lee, Y.-J., Yun, J. S., Oh, K.-W., Han, S.-B., Oh, S., Park, M. H. and Hong, J. T. (2016) 'Inhibition of p38 Pathway-dependent MPTP-induced Dopaminergic Neurodegeneration in Estrogen Receptor Alpha Knockout Mice.', *Hormones and Behavior*, 80, pp. 19–29. doi: 10.1016/j.yhbeh.2016.01.011.

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z. and Dunker, A. K. (2002) 'Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins.', *Journal of Molecular Biology*, 323(3), pp. 573–584. doi: 10.1016/S0022-2836(02)00969-5.

Inestrosa, N. C. and Varela-Nallar, L. (2015) 'Wnt Signalling in Neuronal Differentiation and Development.', *Cell and Tissue Research*, 359(1), pp. 215–223. doi: 10.1007/s00441-014-1996-4.

Isgro, T. A. and Schulten, K. (2005) 'Binding Dynamics of Isolated Nucleoporin Repeat Regions to Importin-β.', *Structure*, 13(12), pp. 1869–1879. doi: 10.1016/j.str.2005.09.007.

Iwatsubo, T. (2003) 'Aggregation of α-Synuclein in the Pathogenesis of Parkinson's Disease.', *Journal of Neurology*, 250, pp. 1–1. doi: 10.1007/s00415-003-1303-x.

Jackson-Lewis, V. and Przedborski, S. (2007) 'Protocol for the MPTP Mouse Model of Parkinson's Disease.', *Nature Protocols*, 2(1), pp. 141–151. doi: 10.1038/nprot.2006.342.

Jahandoost, S., Farhanghian, P. and Abbasi, S. (2017) 'The Effects of Sex Protein Receptors and Sex Steroid Hormone Gene Polymorphisms on Breast Cancer Risk.', *Journal of the National Medical Association*, 109(2), pp. 126–138. doi: 10.1016/j.jnma.2017.02.003.

Jaleel, M., Nichols, R. J., Deak, M., Campbell, D. G., Gillardon, F., Knebel, A. and Alessi, D. R. (2007) 'LRRK2 Phosphorylates Moesin at Threonine-558: Characterization of how Parkinson's Disease Mutants Affect Kinase Activity.', *The Biochemical Journal*, 405(2), pp. 307–17. doi: 10.1042/BJ20070209.

Jankovic, J. (2008) 'Parkinson's Disease: Clinical Features and Diagnosis.', *Journal of Neurology, Neurosurgery, and Psychiatry*, 79(4), pp. 368–76. doi: 10.1136/jnnp.2007.131045.

Japrung, D., Dogan, J., Freedman, K. J., Nadzeyka, A., Bauerdick, S., Albrecht, T., Kim, M. J., Jemth, P. and Edel, J. B. (2013) 'Single-Molecule Studies of Intrinsically Disordered Proteins Using Solid-State Nanopores.', *Analytical Chemistry*, 85(4), pp. 2449–2456. doi: 10.1021/ac3035025.

Jasanoff, A. and Fersht, A. R. (1994) 'Quantitative Determination of Helical Propensities from Trifluoroethanol Titration Curves.', *Biochemistry*, 33, pp. 2129–2135. doi: 10.1021/bi00174a020.

Javadi, Y. and Itzhaki, L. S. (2013) 'Tandem-repeat Proteins: Regularity Plus Modularity Equals Design-ability.', *Current Opinion in Structural Biology*, 23(4), pp. 622–631. doi: 10.1016/j.sbi.2013.06.011.

Javadi, Y. and Main, E. R. G. (2009) 'Exploring the Folding Energy Landscape of a Series of Designed Consensus Tetratricopeptide Repeat Proteins.', *Proceedings of the National Academy of Sciences*, 106(41), pp. 17383–17388. doi: 10.1073/pnas.0907455106.

Jeong, H., Mason, S. P., Barabási, A. L. and Oltvai, Z. N. (2001) 'Lethality and Centrality in Protein Networks.', *Nature*, 411(6833), pp. 41–2. doi: 10.1038/35075138.

Jiang, S.-C., Mei, C., Wang, X.-F. and Zhang, D.-P. (2014) 'A Hub for ABA Signaling to the Nucleus: Significance of a Cytosolic and Nuclear Dual-localized PPR Protein SOAR1 Acting Downstream of Mg-chelatase H Subunit.', *Plant Signaling & Behavior*, 9(11), p. e972899. doi: 10.4161/15592316.2014.972899.

Jin, T. and Liu, L. (2008) 'Minireview: The Wnt Signaling Pathway Effector TCF7L2 and Type 2 Diabetes Mellitus.', *Molecular Endocrinology*, 22(11), pp. 2383–2392. doi: 10.1210/me.2008-0135.

Johansson, K., El-Ahmad, M., Ramaswamy, S., Hjelmqvist, L., Jörnvall, H. and Eklund, H. (1998) 'Structure of Betaine Aldehyde Dehydrogenase at 2.1 Å Resolution.', *Protein Science*, 7(10), pp. 2106–17. doi: 10.1002/pro.5560071007.

Johnson, G. V. W. and Stoothoff, W. H. (2004) 'Tau Phosphorylation in Neuronal Cell Function and Dysfunction.', *Journal of Cell Science*, 117(Pt 24), pp. 5721–9. doi: 10.1242/jcs.01558.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. and Hunter, S. (2014) 'InterProScan 5: Genome-scale Protein Function Classification.', *Bioinformatics*, 30(9), pp. 1236–40. doi: 10.1093/bioinformatics/btu031.

Joslyn, G., Richardson, D. S., White, R. and Alber, T. (1993) 'Dimer Formation by an N-terminal Coiled Coil in the APC Protein.', *Proceedings of the National Academy of Sciences*, 90(23), pp. 11109–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8248216.

Kajava, A. V, Baxa, U. and Steven, A. C. (2010) 'β Arcades: Recurring Motifs in Naturally Occurring and Disease-related Amyloid Fibrils.', *The FASEB Journal*, 24(5), pp. 1311–1319. doi: 10.1096/fj.09-145979.

Kanazawa, M., Ohba, H., Nishiyama, S., Kakiuchi, T. and Tsukada, H. (2017) 'Effect of MPTP on Serotonergic Neuronal Systems and Mitochondrial Complex I Activity in the Living Brain: A PET Study on Conscious Rhesus Monkeys.', *Journal of Nuclear Medicine*, 58(7), pp. 1111–1116. doi: 10.2967/jnumed.116.189159.

Kang, U.-B. and Marto, J. A. (2017) 'Leucine-rich Repeat Kinase 2 and Parkinson's Disease.', *Proteomics*, 17(1–2), p. 1600092. doi: 10.1002/pmic.201600092.

Kapinos, L. E., Schoch, R. L., Wagner, R. S., Schleicher, K. D. and Lim, R. Y. H. (2014) 'Karyopherin-centric Control of Nuclear Pores Based on Molecular Occupancy and Kinetic Analysis of Multivalent Binding with FG Nucleoporins.', *Biophysical Journal*, 106(8), pp. 1751–1762. doi: 10.1016/j.bpj.2014.02.021.

Karush, F. (1950) 'Heterogeneity of the Binding Sites of Bovine Serum Albumin 1.', *Journal of the American Chemical Society*, 72(6), pp. 2705–2713. doi: 10.1021/ja01162a099.

Kelly, L., McDonough, M. A., Coleman, M. L., Ratcliffe, P. J. and Schofield, C. J. (2009) 'Asparagine β-hydroxylation Stabilizes the Ankyrin Repeat Domain Fold.', *Molecular BioSystems*, 5(1), pp. 52–58. doi: 10.1039/B815271C.

Keppel, T. R., Howard, B. A. and Weis, D. D. (2011) 'Mapping Unstructured Regions and Synergistic Folding in Intrinsically Disordered Proteins with Amide H/D Exchange Mass Spectrometry.', *Biochemistry*, 50(40), pp. 8722–8732. doi: 10.1021/bi200875p.

Kiick, K. L., Saxon, E., Tirrell, D. A. and Bertozzi, C. R. (2002) 'Incorporation of Azides into Recombinant Proteins for Chemoselective Modification by the Staudinger Ligation.', *Proceedings of the National Academy of Sciences*, 99(1), pp. 19–24. doi: 10.1073/pnas.012583299.

Kim, W., Kim, M. and Jho, E. (2013) 'Wnt/β-catenin Signalling: from Plasma Membrane to Nucleus.', *Biochemical Journal*, 450(1), pp. 9–21. doi: 10.1042/BJ20121284.

Kimelman, D. and Xu, W. (2006) 'β-catenin Destruction Complex: Insights and Questions from a Structural Perspective.', *Oncogene*, 25(57), pp. 7482–7491. doi: 10.1038/sj.onc.1210055.

Kishida, S., Yamamoto, H., Hino, S., Ikeda, S., Kishida, M. and Kikuchi, A. (1999) 'DIX Domains of Dvl and Axin Are Necessary for Protein Interactions and Their Ability To Regulate β-catenin Stability.', *Molecular and Cellular Biology*, 19(6), pp. 4414–4422. doi: 10.1128/MCB.19.6.4414.

Kitov, P. I. and Bundle, D. R. (2003) 'On the Nature of the Multivalency Effect: A Thermodynamic Model.', *Journal of the American Chemical Society*, 125(52), pp. 16271–16284. doi:

10.1021/ja038223n.

Klein, P., Pawson, T. and Tyers, M. (2003) 'Mathematical Modeling Suggests Cooperative Interactions between a Disordered Polyvalent Ligand and a Single Receptor Site.', *Current Biology*, 13(19), pp. 1669–1678. doi: 10.1016/j.cub.2003.09.027.

Kloss, E., Courtemanche, N. and Barrick, D. (2008) 'Repeat-Protein Folding: New Insights into Origins of Cooperativity, Stability, and Topology.', *Archives of Biochemistry and Biophysics*, 469(1), pp. 83–99. doi: 10.1016/j.abb.2007.08.034.

Kobe, B. and Deisenhofer, J. (1994) 'The Leucine-Rich Repeat: a Versatile Binding Motif.', *Trends in Biochemical Sciences*, 19(10), pp. 415–421. doi: 10.1016/0968-0004(94)90090-6.

Kobe, B. and Deisenhofer, J. (1996) 'Mechanism of Ribonuclease Inhibition by Ribonuclease Inhibitor Protein Based on the Crystal Structure of its Complex with Ribonuclease A.', *Journal of Molecular Biology*, 264(5), pp. 1028–1043. doi: 10.1006/jmbi.1996.0694.

Kobe, B. and Kajava, A. V (2001) 'The Leucine-Rich Repeat as a Protein Recognition Motif.', *Current Opinion in Structural Biology*, 11(6), pp. 725–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11751054 (Accessed: 27 May 2014).

Kobe, B. and Kajava, A. V. (2000) 'When Protein Folding is Simplified to Protein Coiling: the Continuum of Solenoid Protein Structures.', *Trends in Biochemical Sciences*, 25(10), pp. 509–15. doi: 10.1016/S0968-0004(00)01667-4.

Kochańczyk, T., Jakimowicz, P. and Krężel, A. (2013) 'Femtomolar Zn(II) Affinity of Minimal Zinc Hook Peptides – a Promising Small Tag for Protein Engineering.', *Chemical Communications*, 49(13), p. 1312. doi: 10.1039/c2cc38174e.

Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Pluckthun, A. and Grutter, M. G. (2003) 'Designed to be Stable: Crystal Structure of a Consensus Ankyrin Repeat Protein.', *Proceedings of the National Academy of Sciences*, 100(4), pp. 1700–1705. doi: 10.1073/pnas.0337680100.

Kohn, A. D. and Moon, R. T. (2005) 'Wnt and Calcium Signaling: β-catenin-independent Pathways.', *Cell Calcium*, 38(3–4), pp. 439–446. doi: 10.1016/j.ceca.2005.06.022.

Kolb, H. C., Finn, M. G. and Sharpless, K. B. (2001) 'Click Chemistry: Diverse Chemical Function from a Few Good Reactions.', *Angewandte Chemie International Edition*, 40(11), pp. 2004–2021. doi: 10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.

Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B. and Clevers, H. (1997) 'Constitutive Transcriptional Activation by a β-catenin-Tcf Complex in APC-/- Colon Carcinoma.', *Science*, 275(5307), pp. 1784–7. doi: 10.1126/science.275.5307.1784.

Koshland, D. E. (1958) 'Application of a Theory of Enzyme Specificity to Protein Synthesis.', *Proceedings of the National Academy of Sciences*, 44(2), pp. 98–104. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16590179.

Kost, T. A., Condreay, J. P. and Jarvis, D. L. (2005) 'Baculovirus as Versatile Vectors for Protein Expression in Insect and Mammalian Cells.', *Nature Biotechnology*, 23(5), pp. 567–75. doi: 10.1038/nbt1095.

Kozlowski, L. P. and Bujnicki, J. M. (2012) 'MetaDisorder: a Meta-server for the Prediction of Intrinsic Disorder in Proteins.', *BMC Bioinformatics*, 13(1), p. 111. doi: 10.1186/1471-2105-13-111.

Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. and Wright, P. E. (1996) 'Structural Studies of p21Waf1/Cip1/Sdi1 in the Free and Cdk2-bound State: Conformational Disorder Mediates Binding Diversity.', *Proceedings of the National Academy of Sciences*, 93(21), pp. 11504–11509. doi: 10.1073/pnas.93.21.11504.

Krzywda, S., Brzozowski, A. M., Higashitsuji, H., Fujita, J., Welchman, R., Dawson, S., Mayer, R. J. and Wilkinson, A. J. (2004) 'The Crystal Structure of Gankyrin, an Oncoprotein Found in Complexes with Cyclin-dependent Kinase 4, a 19 S Proteasomal ATPase Regulator, and the Tumor Suppressors Rb and p53.', *Journal of Biological Chemistry*, 279(2), pp. 1541–1545. doi: 10.1074/jbc.M310265200.

Kumagai, P. S., DeMarco, R. and Lopes, J. L. S. (2017) 'Advantages of Synchrotron Radiation Circular Dichroism Spectroscopy to Study Intrinsically Disordered Proteins.', *European Biophysics Journal*. doi: 10.1007/s00249-017-1202-1.

Kunttas-Tatli, E., Roberts, D. M. and McCartney, B. M. (2014) 'Self-association of the APC Tumor Suppressor is Required for the Assembly, Stability, and Activity of the Wnt Signaling Destruction Complex.', *Molecular Biology of the Cell*, 25(21), pp. 3424–3436. doi: 10.1091/mbc.E14-04-0885.

de la Roche, M., Rutherford, T. J., Gupta, D., Veprintsev, D. B., Saxty, B., Freund, S. M. and Bienz, M. (2012) 'An Intrinsically Labile α-helix Abutting the BCL9-binding Site of β-catenin is Required for its Inhibition by Carnosic Acid.', *Nature Communications*, p. 680. doi: 10.1038/ncomms1680.

Lamb, A. L. and Newcomer, M. E. (1999) 'The Structure of Retinal Dehydrogenase Type II at 2.7 Å Resolution: Implications for Retinal Specificity.', *Biochemistry*, 38(19), pp. 6003–6011. doi:

10.1021/bi9900471.

Larsen, K. and Madsen, L. B. (2009) 'Sequence Conservation between Porcine and Human LRRK2.', *Molecular Biology Reports*, 36(2), pp. 237–243. doi: 10.1007/s11033-007-9172-5.

Lawrence, A.-M. and Besir, H. (2009) 'Staining of Proteins in Gels with Coomassie G-250 without Organic Solvent and Acetic Acid.', *Journal of Visualized Experiments*, (30). doi: 10.3791/1350.

Leahy, D. J., Hendrickson, W. A., Aukhil, I. and Erickson, H. P. (1992) 'Structure of a Fibronectin Type III Domain from Tenascin Phased by MAD Analysis of the Selenomethionyl Protein.', *Science*, 258(5084), pp. 987–91. doi: 10.1126/science.1279805.

Leibly, D. J., Nguyen, T. N., Kao, L. T., Hewitt, S. N., Barrett, L. K. and Van Voorhis, W. C. (2012) 'Stabilizing Additives Added during Cell Lysis Aid in the Solubilization of Recombinant Proteins.', *PLoS ONE*, 7(12), p. e52482. doi: 10.1371/journal.pone.0052482.

Lesage, S. and Brice, A. (2009) 'Parkinson's Disease: from Monogenic Forms to Genetic Susceptibility Factors.', *Human Molecular Genetics*, 18(R1), pp. R48-59. doi: 10.1093/hmg/ddp012.

Levchenko, A. (2003) 'Allovalency: A Case of Molecular Entanglement.', *Current Biology*, 13(22), pp. R876–R878. doi: 10.1016/j.cub.2003.10.049.

Levine, Z. A., Larini, L., LaPointe, N. E., Feinstein, S. C. and Shea, J.-E. (2015) 'Regulation and Aggregation of Intrinsically Disordered Peptides.', *Proceedings of the National Academy of Sciences*, 112(9), pp. 2758–2763. doi: 10.1073/pnas.1418155112.

Li, D. and Roberts, R. (2001) 'Human Genome and Diseases: WD-repeat Proteins: Structure Characteristics, Biological Function, and their Involvement in Human Diseases.', *Cellular and Molecular Life Sciences*, 58(14), pp. 2085–2097. doi: 10.1007/PL00000838.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V, King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon, B. T. and Rosen, M. K. (2012) 'Phase Transitions in the Assembly of Multivalent Signalling Proteins.', *Nature*, 483(7389), pp. 336–340. doi: 10.1038/nature10879.

Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P. and Yang, B. (2011) 'TAL Nucleases (TALNs): Hybrid Proteins Composed of TAL Effectors and FokI DNA-cleavage Domain.', *Nucleic Acids Research*, 39(1), pp. 359–372. doi: 10.1093/nar/gkq704.

Li, V. S. W., Ng, S. S., Boersema, P. J., Low, T. Y., Karthaus, W. R., Gerlach, J. P., Mohammed, S., Heck,

A. J. R., Maurice, M. M., Mahmoudi, T. and Clevers, H. (2012) 'Wnt Signaling Through Inhibition of β-catenin Degradation in an Intact Axin1 Complex.', *Cell*, 149(6), pp. 1245–56. doi: 10.1016/j.cell.2012.05.002.

Lima, M. M. S., Andersen, M. L., Reksidler, A. B., Vital, M. A. B. F. and Tufik, S. (2007) 'The Role of the Substantia Nigra Pars Compacta in Regulating Sleep Patterns in Rats.', *PloS ONE*, 2(6), p. e513. doi: 10.1371/journal.pone.0000513.

Lin, X., Parisiadou, L., Gu, X.-L., Wang, L., Shim, H., Sun, L., Xie, C., Long, C.-X., Yang, W.-J., Ding, J., Chen, Z. Z., Gallant, P. E., Tao-Cheng, J.-H., Rudow, G., Troncoso, J. C., Liu, Z., Li, Z. and Cai, H. (2009) 'Leucine-Rich Repeat Kinase 2 Regulates the Progression of Neuropathology Induced by Parkinson's-Disease-Related Mutant Alpha-Synuclein.', *Neuron*, 64(6), pp. 807–27. doi: 10.1016/j.neuron.2009.11.006.

Linderstrøm-Lang, K. and Schellman, J. A. (1954) 'The Structure of Insulin as Compared to that of Sanger's A-chain.', *Biochimica et Biophysica Acta*, 15(1), pp. 156–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13198960.

Liu, J., Xing, Y., Hinds, T. R., Zheng, J. and Xu, W. (2006) 'The Third 20 Amino Acid Repeat is the Tightest Binding Site of APC for β-catenin.', *Journal of Molecular Biology*, 360(1), pp. 133–44. doi: 10.1016/j.jmb.2006.04.064.

Liu, W., Dong, X., Mai, M., Seelan, R. S., Taniguchi, K., Krishnadath, K. K., Halling, K. C., Cunningham, J. M., Qian, C., Christensen, E., Roche, P. C., Smith, D. I. and Thibodeau, S. N. (2000) 'Mutations in AXIN2 cause Colorectal Cancer with Defective Mismatch Repair by Activating β-catenin/TCF Signalling.', *Nature Genetics*, 26(2), pp. 146–147. doi: 10.1038/79859.

Liu, Z.-J., Sun, Y.-J., Rose, J., Chung, Y.-J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. and Wang, B.-C. (1997) 'The First Structure of an Aldehyde Dehydrogenase Reveals Novel Interactions between NAD and the Rossmann Fold.', *Nature Structural Biology*, 4(4), pp. 317–326. doi: 10.1038/nsb0497-317.

Ljungberg, T., Apicella, P. and Schultz, W. (1992) 'Responses of Monkey Dopamine Neurons during Learning of Behavioral Reactions.', *Journal of Neurophysiology*, 67(1), pp. 145–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1552316 (Accessed: 22 June 2014).

Locasale, J. W. (2008) 'Allovalency Revisited: An Analysis of Multisite Phosphorylation and Substrate Rebinding.', *The Journal of Chemical Physics*, 128(11), p. 115106. doi: 10.1063/1.2841124.

Logan, C. Y. and Nusse, R. (2004) 'The Wnt Signalling Pathway in Development and Disease.', *Annual Review of Cell and Developmental Biology*, 20(1), pp. 781–810. doi: 10.1146/annurev.cellbio.20.010403.113126.

Löw, C., Homeyer, N., Weininger, U., Sticht, H. and Balbach, J. (2009) 'Conformational Switch upon Phosphorylation: Human CDK Inhibitor p19 INK4d between the Native and Partially Folded State.', *ACS Chemical Biology*, 4(1), pp. 53–63. doi: 10.1021/cb800219m.

Lowe, A. R. and Itzhaki, L. S. (2007) 'Rational Redesign of the Folding Pathway of a Modular Protein.', *Proceedings of the National Academy of Sciences*, 104(8), pp. 2679–2684. doi: 10.1073/pnas.0604653104.

Lowe, E. D., Tews, I., Cheng, K. Y., Brown, N. R., Gul, S., Noble, M. E. M., Gamblin, S. J. and Johnson, L. N. (2002) 'Specificity Determinants of Recruitment Peptides Bound to Phospho-CDK2/cyclin A.', *Biochemistry*, 41(52), pp. 15625–34. doi: 10.1021/bi0268910.

Lux, S. E., John, K. M. and Bennett, V. (1990) 'Analysis of cDNA for Human Erythrocyte Ankyrin Indicates a Repeated Structure with Homology to Tissue-differentiation and Cell-cycle Control Proteins.', *Nature*, 344(6261), pp. 36–42. doi: 10.1038/344036a0.

MacDonald, B. T., Tamai, K. and He, X. (2009) 'Wnt/Beta-Catenin Signaling: Components, Mechanisms, and Diseases.', *Developmental Cell*, 17(1), pp. 9–26. doi: 10.1016/j.devcel.2009.06.016.

Mani, M., Carrasco, D. E., Zhang, Y., Takada, K., Gatt, M. E., Dutta-Simmons, J., Ikeda, H., Diaz-Griffero, F., Pena-Cruz, V., Bertagnolli, M., Myeroff, L. L., Markowitz, S. D., Anderson, K. C. and Carrasco, D. R. (2009) 'BCL9 Promotes Tumor Progression by Conferring Enhanced Proliferative, Metastatic, and Angiogenic Properties to Cancer Cells.', *Cancer Research*, 69(19), pp. 7577–7586. doi: 10.1158/0008-5472.CAN-09-0773.

Marcotte, E. M., Pellegrini, M., Yeates, T. O. and Eisenberg, D. (1999) 'A Census of Protein Repeats.', *Journal of Molecular Biology*, 293(1), pp. 151–160. doi: 10.1006/jmbi.1999.3136.

Matell, M. S. and Meck, W. H. (2000) 'Neuropsychological Mechanisms of Interval Timing Behavior.', *BioEssays*, 22(1), pp. 94–103. doi: 10.1002/(SICI)1521-1878(200001)22:1<94::AID-BIES14>3.0.CO;2-E.

Matouschek, A., Kellis, J. T., Serrano, L. and Fersht, A. R. (1989) 'Mapping the Transition State and Pathway of Protein Folding by Protein Engineering.', *Nature*, 340(6229), pp. 122–126. doi: 10.1038/340122a0.

McCartney, B. M. and Näthke, I. S. (2008) 'Cell Regulation by the APC Protein: APC as Master

Regulator of Epithelia.', *Current Opinion in Cell Biology*, 20(2), pp. 186–193. doi: 10.1016/j.ceb.2008.02.001.

McNulty, B. C., Young, G. B. and Pielak, G. J. (2006) 'Macromolecular Crowding in the Escherichia coli Periplasm Maintains α-Synuclein Disorder.', *Journal of Molecular Biology*, 355(5), pp. 893–897. doi: 10.1016/j.jmb.2005.11.033.

Meek, D. W. (2015) 'Regulation of the p53 Response and its Relationship to Cancer.', *Biochemical Journal*, 469(3), pp. 325–346. doi: 10.1042/BJ20150517.

Mello, C. C. and Barrick, D. (2004) 'An Experimentally Determined Protein Folding Energy Landscape.', *Proceedings of the National Academy of Sciences*, 101(39), pp. 14102–14107. doi: 10.1073/pnas.0403386101.

Merrifield, R. B. (1963) 'Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide.', *Journal of the American Chemical Society*, 85(14), pp. 2149–2154. doi: 10.1021/ja00897a025.

Michaely, P. (2002) 'Crystal Structure of a 12 ANK Repeat Stack from Human AnkyrinR.', *The EMBO Journal*, 21(23), pp. 6387–6396. doi: 10.1093/emboj/cdf651.

Milles, S., Mercadante, D., Aramburu, I. V., Jensen, M. R., Banterle, N., Koehler, C., Tyagi, S., Clarke, J., Shammas, S. L., Blackledge, M., Gräter, F. and Lemke, E. A. (2015) 'Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors.', *Cell*, 163(3), pp. 734–745. doi: 10.1016/j.cell.2015.09.047.

Mills, R. D., Mulhern, T. D., Cheng, H.-C. and Culvenor, J. G. (2012) 'Analysis of LRRK2 Accessory Repeat Domains: Prediction of Repeat Length, Number and Sites of Parkinson's Disease Mutations.', *Biochemical Society Transactions*, 40(5), pp. 1086–9. doi: 10.1042/BST20120088.

Minde, D. P., Anvarian, Z., Rüdiger, S. G. and Maurice, M. M. (2011) 'Messing up Disorder: How do Missense Mutations in the Tumor Suppressor Protein APC lead to Cancer?', *Molecular Cancer*, 10, p. 101. doi: 10.1186/1476-4598-10-101.

Miroux, B. and Walker, J. E. (1996) 'Over-production of Proteins in Escherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels.', *Journal of Molecular Biology*, 260(3), pp. 289–98. doi: 10.1006/jmbi.1996.0399.

Mitrea, D. M., Grace, C. R., Buljan, M., Yun, M.-K., Pytel, N. J., Satumba, J., Nourse, A., Park, C.-G., Madan Babu, M., White, S. W. and Kriwacki, R. W. (2014) 'Structural Polymorphism in the N-terminal Oligomerization Domain of NPM1.', *Proceedings of the National Academy of Sciences*, 111(12), pp.

4466–4471. doi: 10.1073/pnas.1321007111.

Mittag, T., Kay, L. E. and Forman-Kay, J. D. (2009) 'Protein Dynamics and Conformational Disorder in Molecular Recognition.', *Journal of Molecular Recognition*, 23(2), pp. 105–16. doi: 10.1002/jmr.961.

Mittag, T., Orlicky, S., Choy, W.-Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M. and Forman-Kay, J. D. (2008) 'Dynamic Equilibrium Engagement of a Polyvalent Ligand with a Single-site Receptor.', *Proceedings of the National Academy of Sciences*, 105(46), pp. 17772–7. doi: 10.1073/pnas.0809222105.

Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K. and Uversky, V. N. (2006) 'Analysis of Molecular Recognition Features (MoRFs).', *Journal of Molecular Biology*, 362(5), pp. 1043–1059. doi: 10.1016/j.jmb.2006.07.087.

Monod, J., Wyman, J. and Changeux, J.-P. (1965) 'On the Nature of Allosteric Transitions: A Plausible Model.', *Journal of Molecular Biology*, 12(1), pp. 88–118. doi: 10.1016/S0022-2836(65)80285-6.

Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. and Baker, E. N. (1998) 'Sheep Liver Cytosolic Aldehyde Dehydrogenase: the Structure Reveals the Basis for the Retinal Specificity of Class 1 Aldehyde Dehydrogenases.', *Structure*, 6(12), pp. 1541–1551. doi: 10.1016/S0969-2126(98)00152-X.

Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. and Kinzler, K. W. (1997) 'Activation of β-catenin-Tcf Signaling in Colon Cancer by Mutations in β-catenin or APC.', *Science*, 275(5307), pp. 1787–90. doi: 10.1126/science.275.5307.1787.

Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. and Peng, Z.-Y. (2004) 'The Ankyrin Repeat as Molecular Architecture for Protein Recognition.', *Protein Science*, 13(6), pp. 1435–1448. doi: 10.1110/ps.03554604.

Mosavi, L. K., Minor, D. L. and Peng, Z.-Y. (2002) 'Consensus-Derived Structural Determinants of the Ankyrin Repeat Motif.', *Proceedings of the National Academy of Sciences*, 99(25), pp. 16029–34. doi: 10.1073/pnas.252537899.

Mosimann, C., Hausmann, G. and Basler, K. (2009) 'β-Catenin hits Chromatin: Regulation of Wnt Target Gene Activation.', *Nature Reviews Molecular Cell Biology*, 10(4), pp. 276–286. doi: 10.1038/nrm2654.

Moussaud, S., Jones, D. R., Moussaud-Lamodière, E. L., Delenclos, M., Ross, O. A. and McLean, P. J. (2014) 'Alpha-synuclein and Tau: Teammates in Neurodegeneration?', *Molecular*

Neurodegeneration, 9(1), p. 43. doi: 10.1186/1750-1326-9-43.

Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, Sanchez, R., Zeleznik-Le, N. J., Ronai, Z. and Zhou, M.-M. (2004) 'Structural Mechanism of the Bromodomain of the Coactivator CBP in p53 Transcriptional Activation.', *Molecular Cell*, 13(2), pp. 251–263. doi: 10.1016/S1097-2765(03)00528-8.

Murray, N. E., Bruce, S. A. and Murray, K. (1979) 'Molecular Cloning of the DNA Ligase Gene from Bacteriophage T4.', *Journal of Molecular Biology*, 132(3), pp. 493–505. doi: 10.1016/0022-2836(79)90271-7.

Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F. B., Mendenhall, M. D., Sicheri, F., Pawson, T. and Tyers, M. (2001) 'Multisite Phosphorylation of a CDK Inhibitor sets a Threshold for the Onset of DNA Replication.', *Nature*, 414(6863), pp. 514–521. doi: 10.1038/35107009.

NEB (no date) *Excess Nucleobase Requirements of Different Restriction Enzymes.* Available at: https://www.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments#chart-H.

Neer, E. J., Schmidt, C. J., Nambudripad, R. and Smith, T. F. (1994) 'The Ancient Regulatory-protein Family of WD-repeat Proteins.', *Nature*, 371(6495), pp. 297–300. doi: 10.1038/371297a0.

Nichols, R. J., Dzamko, N., Morrice, N. A., Campbell, D. G., Deak, M., Ordureau, A., Macartney, T., Tong, Y., Shen, J., Prescott, A. R. and Alessi, D. R. (2010) '14-3-3 Binding to LRRK2 is Disrupted by Multiple Parkinson's Disease-associated Mutations and Regulates Cytoplasmic Localization.', *The Biochemical Journal*, 430(3), pp. 393–404.

Niehrs, C. and Acebron, S. P. (2010) 'Wnt Signaling: Multivesicular Bodies Hold GSK3 Captive.', *Cell*, 143(7), pp. 1044–1046. doi: 10.1016/j.cell.2010.12.003.

Niu, J., Yu, M., Wang, C. and Xu, Z. (2012) 'Leucine-rich Repeat Kinase 2 Disturbs Mitochondrial Dynamics via Dynamin-like Protein.', *Journal of Neurochemistry*, 122(3), pp. 650–658. doi: 10.1111/j.1471-4159.2012.07809.x.

Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D., Pawson, T., Forman-Kay, J. and Baldwin, A. (2015) 'Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles.', *Molecular Cell*, 57(5), pp. 936–947. doi: 10.1016/j.molcel.2015.01.013.

Noutsou, M., Duarte, A. M. S., Anvarian, Z., Didenko, T., Minde, D. P., Kuper, I., de Ridder, I.,

Oikonomou, C., Friedler, A., Boelens, R., Rüdiger, S. G. D. and Maurice, M. M. (2011) 'Critical Scaffolding Regions of the Tumor Suppressor Axin1 are Natively Unfolded.', *Journal of Molecular Biology*, 405(3), pp. 773–86. doi: 10.1016/j.jmb.2010.11.013.

Olsen, J. G., Teilum, K. and Kragelund, B. B. (2017) 'Behaviour of Intrinsically Disordered Proteins in Protein-protein Complexes with an Emphasis on Fuzziness.', *Cellular and Molecular Life Sciences*. doi: 10.1007/s00018-017-2560-7.

Omer, C. A., Miller, P. J., Diehl, R. E. and Kral, A. M. (1999) 'Identification of Tcf4 Residues Involved in High-Affinity β-catenin Binding.', *Biochemical and Biophysical Research Communications*, 256(3), pp. 584–590. doi: 10.1006/bbrc.1999.0379.

OpenWetWare (2013) *Round-the-Horn Site-Directed Mutagensis*. Available at: http://openwetware.org/wiki/'Round-the-horn_site-directed_mutagenesis.

Ott, W., Jobst, M. A., Schoeler, C., Gaub, H. E. and Nash, M. A. (2017) 'Single-molecule Force Spectroscopy on Polyproteins and Receptor–ligand Complexes: The Current Toolbox.', *Journal of Structural Biology*, 197(1), pp. 3–12. doi: 10.1016/j.jsb.2016.02.011.

Owczarzy, R., You, Y., Moreira, B. G., Manthey, J. A., Huang, L., Behlke, M. A. and Walder, J. A. (2004) 'Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures.', *Biochemistry*, 43(12), pp. 3537–3554. doi: 10.1021/bi034621r.

Ozkan, A. D., Topal, A. E., Dana, A., Guler, M. O. and Tekinay, A. B. (2016) 'Atomic Force Microscopy for the Investigation of Molecular and Cellular Behavior.', *Micron*, 89, pp. 60–76. doi: 10.1016/j.micron.2016.07.011.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) 'How to Measure and Predict the Molar Absorption Coefficient of a Protein.', *Protein Science*, 4(11), pp. 2411–23. doi: 10.1002/pro.5560041120.

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., Hau, J., Martin, O., Kuznetsov, D. and Falquet, L. (2007) 'MyHits: Improvements to an Interactive Resource for Analyzing Protein Sequences.', *Nucleic Acids Research*, 35(SUPPL.2).

Palmer, M. (2014) *WatCut: An On-line Tool for Restriction Analysis, Silent Mutation Scanning, and SNP-RFLP Analysis.* Available at: http://watcut.uwaterloo.ca/template.php.

Parish, C. L., Castelo-Branco, G., Rawal, N., Tonnesen, J., Sorensen, A. T., Salto, C., Kokaia, M., Lindvall, O. and Arenas, E. (2008) 'Wnt5a-treated Midbrain Neural Stem Cells Improve Dopamine Cell

Replacement Therapy in Parkinsonian Mice.', *Journal of Clinical Investigation*, 118(1), pp. 149–160. doi: 10.1172/JCI32273.

Parisiadou, L. and Cai, H. (2010) 'LRRK2 Function on Actin and Microtubule Dynamics in Parkinson Disease.', *Communicative & Integrative Biology*, 3(5), pp. 396–400. doi: 10.4161/cib.3.5.12286.

Parisiadou, L., Xie, C., Cho, H. J., Lin, X., Gu, X.-L., Long, C.-X., Lobbestael, E., Baekelandt, V., Taymans, J.-M., Sun, L. and Cai, H. (2009) 'Phosphorylation of Ezrin/Radixin/Moesin Proteins by LRRK2 Promotes the Rearrangement of Actin Cytoskeleton in Neuronal Morphogenesis.', *The Journal of Neuroscience*, 29(44), pp. 13971–80. doi: 10.1523/JNEUROSCI.3799-09.2009.

Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., de Castro, E., Baratin, D., Cuche, B. A., Bougueleret, L., Poux, S., Redaschi, N., Xenarios, I. and Bridge, A. (2013) 'HAMAP in 2013, New Developments in the Protein Family Classification and Annotation System.', *Nucleic Acids Research*, 41(D1), pp. D584–D589. doi: 10.1093/nar/gks1157.

Peifer, M., Berg, S. and Reynolds, A. (1994) 'A Repeating Amino Acid Motif Shared by Proteins with Diverse Cellular Roles.', *Cell*, 76(5), pp. 789–791. doi: 10.1016/0092-8674(94)90353-0.

Perilla, J. R., Zhao, G., Lu, M., Ning, J., Hou, G., Byeon, I.-J. L., Gronenborn, A. M., Polenova, T. and Zhang, P. (2017) 'CryoEM Structure Refinement by Integrating NMR Chemical Shifts with Molecular Dynamics Simulations.', *The Journal of Physical Chemistry B*, p. acs.jpcb.6b13105. doi: 10.1021/acs.jpcb.6b13105.

Peti, W., Page, R., Boura, E. and Różycki, B. (2018) 'Structures of Dynamic Protein Complexes: Hybrid Techniques to Study MAP Kinase Complexes and the ESCRT System.', in Ghose, R. (ed.) *Protein NMR: Methods and Protocols*. New York, NY: Springer New York, pp. 375–389. doi: 10.1007/978-1-4939-7386-6_17.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) 'UCSF Chimera - A Visualization System for Exploratory Research and Analysis.', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. doi: 10.1002/jcc.20084.

Piccoli, G., Condliffe, S. B., Bauer, M., Giesert, F., Boldt, K., De Astis, S., Meixner, A., Sarioglu, H., Vogt-Weisenhorn, D. M., Wurst, W., Gloeckner, C. J., Matteoli, M., Sala, C. and Ueffing, M. (2011) 'LRRK2 Controls Synaptic Vesicle Storage and Mobilization within the Recycling Pool.', *Journal of Neuroscience*, 31(6), pp. 2225–2237. doi: 10.1523/JNEUROSCI.3730-10.2011.

Piccoli, G., Onofri, F., Cirnaru, M. D., Kaiser, C. J. O., Jagtap, P., Kastenmuller, A., Pischedda, F., Marte,

A., von Zweydorf, F., Vogt, A., Giesert, F., Pan, L., Antonucci, F., Kiel, C., Zhang, M., Weinkauf, S., Sattler, M., Sala, C., Matteoli, M., Ueffing, M. and Gloeckner, C. J. (2014) 'Leucine-rich Repeat Kinase 2 Binds to Neuronal Vesicles through Protein Interactions Mediated by its C-terminal WD40 Domain.', *Molecular and Cellular Biology*, 34(12), pp. 2147–2161. doi: 10.1128/MCB.00914-13.

Pioli, E. Y., Meissner, W., Sohr, R., Gross, C. E., Bezard, E. and Bioulac, B. H. (2008) 'Differential Behavioral Effects of Partial Bilateral Lesions of Ventral Tegmental Area or Substantia Nigra Pars Compacta in Rats.', *Neuroscience*, 153(4), pp. 1213–24. doi: 10.1016/j.neuroscience.2008.01.084.

Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C. J., Aspromonte, M. C., Davey, N. E., Davidović, R., Dosztányi, Z., Elofsson, A., Gasparini, A., Hatos, A., Kajava, A. V, Kalmar, L., Leonardi, E., Lazar, T., Macedo-Ribeiro, S., Macossay-Castillo, M., Meszaros, A., Minervini, G., Murvai, N., Pujols, J., Roche, D. B., Salladini, E., Schad, E., Schramm, A., Szabo, B., Tantos, A., Tonello, F., Tsirigos, K. D., Veljković, N., Ventura, S., Vranken, W., Warholm, P., Uversky, V. N., Dunker, A. K., Longhi, S., Tompa, P. and Tosatto, S. C. E. (2017) 'DisProt 7.0: a Major Update of the Database of Disordered Proteins.', *Nucleic Acids Research*, 45(D1), pp. D219–D227. doi: 10.1093/nar/gkw1056.

Plückthun, A. (2015) 'Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy.', *Annual Review of Pharmacology and Toxicology*, 55(1), pp. 489–511. doi: 10.1146/annurev-pharmtox-010611-134654.

Pokutta, S. and Weis, W. I. (2007) 'Structure and Mechanism of Cadherins and Catenins in Cell-Cell Contacts.', *Annual Review of Cell and Developmental Biology*, 23(1), pp. 237–261. doi: 10.1146/annurev.cellbio.22.010305.104241.

Polakis, P. (2007) 'The Many Ways of Wnt in Cancer.', *Current Opinion in Genetics & Development*, 17(1), pp. 45–51. doi: 10.1016/j.gde.2006.12.007.

Popken, P., Ghavami, A., Onck, P. R., Poolman, B. and Veenhoff, L. M. (2015) 'Size-dependent Leak of Soluble and Membrane Proteins through the Yeast Nuclear Pore Complex.', *Molecular Biology of the Cell*, 26(7), pp. 1386–1394. doi: 10.1091/mbc.E14-07-1175.

Port, F. and Basler, K. (2010) 'Wnt Trafficking: New Insights into Wnt Maturation, Secretion and Spreading.', *Traffic*, 11(10), pp. 1265–1271. doi: 10.1111/j.1600-0854.2010.01076.x.

Poy, F., Lepourcelet, M., Shivdasani, R. A. and Eck, M. J. (2001) 'Structure of a Human Tcf4-β-catenin Complex.', *Nature Structural Biology*, 8(12), pp. 1053–7. doi: 10.1038/nsb720.

Punchihewa, C., Ferreira, A. M., Cassell, R., Rodrigues, P. and Fujii, N. (2009) 'Sequence Requirement

and Subtype Specificity in the High-affinity Interaction between Human Frizzled and Dishevelled Proteins.', *Protein Science*, 18(5), pp. 994–1002. doi: 10.1002/pro.109.

Qi, Y.-K., Chang, H.-N., Pan, K.-M., Tian, C.-L. and Zheng, J.-S. (2015) 'Total Chemical Synthesis of the Site-selective Azide-labeled [I66A]HIV-1 Protease.', *Chemical Communications*, 51(78), pp. 14632–14635. doi: 10.1039/C5CC04846J.

Qing, H., Wong, W., McGeer, E. G. and McGeer, P. L. (2009) 'LRRK2 Phosphorylates Alpha Synuclein at Serine 129: Parkinson Disease Implications.', *Biochemical and Biophysical Research Communications*, 387(1), pp. 149–52. doi: 10.1016/j.bbrc.2009.06.142.

Rani, P., Baruah, A. and Biswas, P. (2014) 'Does Lack of Secondary Structure Imply Intrinsic Disorder in Proteins? A Sequence Analysis.', *Biochimica et Biophysica Acta*, 1844(10), pp. 1827–1834. doi: 10.1016/j.bbapap.2014.07.020.

Recchia, A., Debetto, P., Negro, A., Guidolin, D., Skaper, S. D. and Giusti, P. (2004) 'α-Synuclein and Parkinson's Disease.', *FASEB Journal*, 18(6), pp. 617–26. doi: 10.1096/fj.03-0338rev.

Reichen, C., Hansen, S. and Plückthun, A. (2014) 'Modular Peptide Binding: From a Comparison of Natural Binders to Designed Armadillo Repeat Proteins.', *Journal of Structural Biology*, 185(2), pp. 147–162. doi: 10.1016/j.jsb.2013.07.012.

Rina, M., Pozidis, C., Mavromatis, K., Tzanodaskalaki, M., Kokkinidis, M. and Bouriotis, V. (2000) 'Alkaline Phosphatase from the Antarctic Strain TAB5. Properties and Psychrophilic Adaptations.', *European Journal of Biochemistry / FEBS*, 267(4), pp. 1230–8. doi: 10.1046/j.1432-1327.2000.01127.x.

Rizzo, G., Martinelli, P., Manners, D., Scaglione, C., Tonon, C., Cortelli, P., Malucelli, E., Capellari, S., Testa, C., Parchi, P., Montagna, P., Barbiroli, B. and Lodi, R. (2008) 'Diffusion-weighted Brain Imaging Study of Patients with Clinical Diagnosis of Corticobasal Degeneration, Progressive Supranuclear Palsy and Parkinson's Disease.', *Brain*, 131(10), pp. 2690–2700. doi: 10.1093/brain/awn195.

Rogers, J. M., Steward, A. and Clarke, J. (2013) 'Folding and Binding of an Intrinsically Disordered Protein: Fast, but Not "Diffusion-limited".', *Journal of the American Chemical Society*, 135(4), pp. 1415–1422. doi: 10.1021/ja309527h.

Rogers, J. M., Wong, C. T. and Clarke, J. (2014) 'Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure.', *Journal of the American Chemical Society*, 136(14), pp. 5197–5200. doi: 10.1021/ja4125065.

Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., Costanzo, L. Di, Duarte, J. M., Dutta, S., Feng, Z., Green, R. K., Goodsell, D. S., Hudson, B., Kalro, T., Lowe, R., Peisach, E., Randle, C., Rose, A. S., Shao, C., Tao, Y.-P., Valasatava, Y., Voigt, M., Westbrook, J. D., Woo, J., Yang, H., Young, J. Y., Zardecki, C., Berman, H. M. and Burley, S. K. (2017) 'The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information.', *Nucleic Acids Research*, 45(D1), pp. D271–D281. doi: 10.1093/nar/gkw1000.

Rossi, A. M. and Taylor, C. W. (2011) 'Analysis of Protein-ligand Interactions by Fluorescence Polarization.', *Nature Protocols*, 6(3), pp. 365–387. doi: 10.1038/nprot.2011.305.

Rout, M. P., Aitchison, J. D., Magnasco, M. O. and Chait, B. T. (2003) 'Virtual Gating and Nuclear Transport: the Hole Picture.', *Trends in Cell Biology*, 13(12), pp. 622–8. doi: 10.1016/j.tcb.2003.10.007.

Rowling, P. J. E., Sivertsson, E. M., Perez-Riba, A., Main, E. R. G. and Itzhaki, L. S. (2015) 'Dissecting and Reprogramming the Folding and Assembly of Tandem-repeat Proteins.', *Biochemical Society Transactions*, 43(5), pp. 881–888. doi: 10.1042/BST20150099.

Russo, I., Bubacco, L. and Greggio, E. (2014) 'LRRK2 and Neuroinflammation: Partners in Crime in Parkinson's Disease?', *Journal of Neuroinflammation*, 11(1), p. 52. doi: 10.1186/1742-2094-11-52.

Rustandi, R. R., Baldisseri, D. M. and Weber, D. J. (2000) 'Structure of the Negative Regulatory Domain of p53 Bound to S100B(ββ).', *Nature Structural Biology*, 7(7), pp. 570–4. doi: 10.1038/76797.

Sakiyama, Y., Panatala, R. and Lim, R. Y. H. (2017) 'Structural Dynamics of the Nuclear Pore Complex.', *Seminars in Cell & Developmental Biology*. doi: 10.1016/j.semcdb.2017.05.021.

Sampietro, J., Dahlberg, C. L., Cho, U. S., Hinds, T. R., Kimelman, D. and Xu, W. (2006) 'Crystal Structure of a β-catenin/BCL9/Tcf4 Complex.', *Molecular Cell*, 24(2), pp. 293–300. doi: 10.1016/j.molcel.2006.09.001.

SantaLucia, J. (1998) 'A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics.', *Proceedings of the National Academy of Sciences*, 95(4), pp. 1460–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9465037.

Santos, D., Esteves, A. R., Silva, D. F., Januário, C. and Cardoso, S. M. (2015) 'The Impact of Mitochondrial Fusion and Fission Modulation in Sporadic Parkinson's Disease.', *Molecular Neurobiology*, 52(1), pp. 573–586. doi: 10.1007/s12035-014-8893-4.

Sato, H., Kato, T. and Arawaka, S. (2013) 'The Role of Ser129 Phosphorylation of α-synuclein in

Neurodegeneration of Parkinson's Disease: a Review of in vivo Models.', *Reviews in the Neurosciences*, 24(2), pp. 115–23. doi: 10.1515/revneuro-2012-0071.

Schoch, R. L., Kapinos, L. E. and Lim, R. Y. H. (2012) 'Nuclear Transport Receptor Binding Avidity Triggers a Self-healing Collapse Transition in FG-nucleoporin Molecular Brushes.', *Proceedings of the National Academy of Sciences*, 109(42), pp. 16911–16916. doi: 10.1073/pnas.1208440109.

Schwarz-Romond, T., Metcalfe, C. and Bienz, M. (2007) 'Dynamic Recruitment of Axin by Dishevelled Protein Assemblies.', *Journal of Cell Science*, 120(14), pp. 2402–2412. doi: 10.1242/jcs.002956.

Schweers, O., Schönbrunn-Hanebeck, E., Marx, A. and Mandelkow, E. (1994) 'Structural Studies of Tau Protein and Alzheimer Paired Helical Filaments Show no Evidence for β-structure.', *The Journal of Biological Chemistry*, 269(39), pp. 24290–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7929085.

Seifert, J. R. K. and Mlodzik, M. (2007) 'Frizzled/PCP Signalling: a Conserved Mechanism Regulating Cell Polarity and Directed Motility.', *Nature Reviews Genetics*, 8(2), pp. 126–138. doi: 10.1038/nrg2042.

Sharma, R., Raduly, Z., Miskei, M. and Fuxreiter, M. (2015) 'Fuzzy Complexes: Specific Binding Without Complete Folding.', *FEBS Letters*, 589(19 Pt A), pp. 2533–42. doi: 10.1016/j.febslet.2015.07.022.

Sheinerman, F., Norel, R. and Honig, B. (2000) 'Electrostatic Aspects of Protein–protein Interactions.', *Current Opinion in Structural Biology*, 10(2), pp. 153–159. doi: 10.1016/S0959-440X(00)00065-8.

Sibille, N. and Bernadó, P. (2012) 'Structural Characterization of Intrinsically Disordered Proteins by the Combined Use of NMR and SAXS.', *Biochemical Society Transactions*, 40(5), pp. 955–962. doi: 10.1042/BST20120149.

Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z. and Dunker, A. K. (2007) 'DisProt: the Database of Disordered Proteins.', *Nucleic Acids Research*, 35(Database), pp. D786–D793. doi: 10.1093/nar/gkl893.

Sigalov, A. B., Kim, W. M., Saline, M. and Stern, L. J. (2008) 'The Intrinsically Disordered Cytoplasmic Domain of the T Cell Receptor ζ Chain Binds to the Nef Protein of Simian Immunodeficiency Virus without a Disorder-to-order Transition?', *Biochemistry*, 47(49), pp. 12942–12944. doi: 10.1021/bi801602p.

Sigalov, A. B., Zhuravleva, A. V and Orekhov, V. Y. (2007) 'Binding of Intrinsically Disordered Proteins is not Necessarily Accompanied by a Structural Transition to a Folded Form.', *Biochimie*, 89(3), pp. 419–421. doi: 10.1016/j.biochi.2006.11.003.

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L. and Xenarios, I. (2013) 'New and Continuing Developments at PROSITE.', *Nucleic Acids Research*, 41(D1), pp. D344–D347. doi: 10.1093/nar/gks1067.

Simon, M., Zangemeister-Wittke, U. and Plückthun, A. (2012) 'Facile Double-functionalization of Designed Ankyrin Repeat Proteins Using Click and Thiol Chemistries.', *Bioconjugate Chemistry*, 23(2), pp. 279–86. doi: 10.1021/bc200591x.

Singh, G. P., Ganapathi, M. and Dash, D. (2006) 'Role of Intrinsic Disorder in Transient Interactions of Hub Proteins.', *Proteins: Structure, Function, and Bioinformatics*, 66(4), pp. 761–765. doi: 10.1002/prot.21281.

Singh, S. M. and Panda, A. K. (2005) 'Solubilization and Refolding of Bacterial Inclusion Body Proteins.', *Journal of Bioscience and Bioengineering*, 99(4), pp. 303–310. doi: 10.1263/jbb.99.303.

Slattery, M. L., Folsom, A. R., Wolff, R., Herrick, J., Caan, B. J. and Potter, J. D. (2008) 'Transcription Factor 7-like 2 Polymorphism and Colon Cancer.', *Cancer Epidemiology Biomarkers & Prevention*, 17(4), pp. 978–982. doi: 10.1158/1055-9965.EPI-07-2687.

Sobolev, V., Eyal, E., Gerzon, S., Potapov, V., Babor, M., Prilusky, J. and Edelman, M. (2005) 'SPACE: a Suite of Tools for Protein Structure Prediction and Analysis Based on Complementarity and Environment.', *Nucleic Acids Research*, 33(Web Server), pp. W39–W43. doi: 10.1093/nar/gki398.

Song, J., Lee, M. S., Carlberg, I., Vener, A. V and Markley, J. L. (2006) 'Micelle-induced Folding of Spinach Thylakoid Soluble Phosphoprotein of 9 kDa and Its Functional Implications.', *Biochemistry*, 45(51), pp. 15633–15643. doi: 10.1021/bi062148m.

Souied, E. H., Devin, F., Mauget-Faÿsse, M., Kolář, P., Wolf-Schnurrbusch, U., Framme, C., Gaucher, D., Querques, G., Stumpp, M. T. and Wolf, S. (2014) 'Treatment of Exudative Age-related Macular Degeneration with a Designed Ankyrin Repeat Protein that Binds Vascular Endothelial Growth Factor: a Phase I/II Study.', *American Journal of Ophthalmology*, 158(4), p. 724–732.e2. doi: 10.1016/j.ajo.2014.05.037.

Sprague, E. R. (2000) 'Structure of the C-terminal Domain of Tup1, a Corepressor of Transcription in Yeast.', *The EMBO Journal*, 19(12), pp. 3016–3027. doi: 10.1093/emboj/19.12.3016.

Stafa, K., Tsika, E., Moser, R., Musso, A., Glauser, L., Jones, A., Biskup, S., Xiong, Y., Bandopadhyay, R., Dawson, V. L., Dawson, T. M. and Moore, D. J. (2014) 'Functional Interaction of Parkinson's Disease-associated LRRK2 with Members of the Dynamin GTPase Superfamily.', *Human Molecular Genetics*, 23(8), pp. 2055–2077. doi: 10.1093/hmg/ddt600.

Stefan, M. I., Edelstein, S. J. and Le Novère, N. (2009) 'Computing Phenomenologic Adair-Klotz Constants from Microscopic MWC Parameters.', *BMC Systems Biology*, 3(1), p. 68. doi: 10.1186/1752-0509-3-68.

Steinmetz, C. G., Xie, P., Weiner, H. and Hurley, T. D. (1997) 'Structure of Mitochondrial Aldehyde Dehydrogenase: the Genetic Component of Ethanol Aversion.', *Structure*, 5(5), pp. 701–711. doi: 10.1016/S0969-2126(97)00224-4.

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R. and Hogenesch, J. B. (2004) 'A Gene Atlas of the Mouse and Human Protein-encoding Transcriptomes.', *Proceedings of the National Academy of Sciences*, 101(16), pp. 6062–6067. doi: 10.1073/pnas.0400782101.

Su, Y.-C. and Qi, X. (2013) 'Inhibition of Excessive Mitochondrial Fission Reduced Aberrant Autophagy and Neuronal Damage caused by LRRK2 G2019S Mutation.', *Human Molecular Genetics*, 22(22), pp. 4545–4561. doi: 10.1093/hmg/ddt301.

Sugase, K., Lansing, J. C., Dyson, H. J. and Wright, P. E. (2007) 'Tailoring Relaxation Dispersion Experiments for Fast-associating Protein Complexes', *Journal of the American Chemical Society*, 129(44), pp. 13406–13407. doi: 10.1021/ja0762238.

Sun, J. and Weis, W. I. (2011) 'Biochemical and Structural Characterization of β-catenin Interactions with Nonphosphorylated and CK2-phosphorylated Lef-1.', *Journal of Molecular Biology*, 405(2), pp. 519–30. doi: 10.1016/j.jmb.2010.11.010.

Sun, N. and Zhao, H. (2013) 'Transcription Activator-like Effector Nucleases (TALENs): A Highly Efficient and Versatile Tool for Genome Editing.', *Biotechnology and Bioengineering*, 110(7), pp. 1811–1821. doi: 10.1002/bit.24890.

Szasz, C., Alexa, A., Toth, K., Rakacs, M., Langowski, J. and Tompa, P. (2011) 'Protein Disorder Prevails under Crowded Conditions.', *Biochemistry*, 50(26), pp. 5834–5844. doi: 10.1021/bi200365j.

Taelman, V. F., Dobrowolski, R., Plouhinec, J.-L., Fuentealba, L. C., Vorwald, P. P., Gumper, I., Sabatini, D. D. and De Robertis, E. M. (2010) 'Wnt Signaling Requires Sequestration of Glycogen

Synthase Kinase 3 inside Multivesicular Endosomes.', *Cell*, 143(7), pp. 1136–1148. doi: 10.1016/j.cell.2010.11.034.

Takada, K., Zhu, D., Bird, G. H., Sukhdeo, K., Zhao, J.-J., Mani, M., Lemieux, M., Carrasco, D. E., Ryan, J., Horst, D., Fulciniti, M., Munshi, N. C., Xu, W., Kung, A. L., Shivdasani, R. A., Walensky, L. D. and Carrasco, D. R. (2012) 'Targeted Disruption of the BCL9/β-catenin Complex Inhibits Oncogenic Wnt Signaling.', *Science Translational Medicine*, p. 148ra117-148ra117. doi: 10.1126/scitranslmed.3003808.

Takahashi, N., Takahashi, Y. and Putnam, F. W. (1985) 'Periodicity of Leucine and Tandem Repetition of a 24-amino Acid Segment in the Primary Structure of Leucine-rich α 2-glycoprotein of Human Serum.', *Proceedings of the National Academy of Sciences*, 82(7), pp. 1906–1910. doi: 10.1073/pnas.82.7.1906.

Tamai, K., Zeng, X., Liu, C., Zhang, X., Harada, Y., Chang, Z. and He, X. (2004) 'A Mechanism for Wnt Coreceptor Activation.', *Molecular Cell*, 13(1), pp. 149–156. doi: 10.1016/S1097-2765(03)00484-2.

Tan, E.-K. and Jankovic, J. (2006) 'Genetic Testing in Parkinson Disease: Promises and Pitfalls.', *Archives of Neurology*, 63(9), pp. 1232–7. doi: 10.1001/archneur.63.9.1232.

Tang, K. S., Fersht, A. R. and Itzhaki, L. S. (2003) 'Sequential Unfolding of Ankyrin Repeats in Tumor Suppressor p16.', *Structure*, 11(1), pp. 67–73. doi: 10.1016/S0969-2126(02)00929-2.

Tang, W., Dodge, M., Gundapaneni, D., Michnoff, C., Roth, M. and Lum, L. (2008) 'A Genome-wide RNAi Screen for Wnt/β-catenin Pathway Components Identifies Unexpected Roles for TCF Transcription Factors in Cancer.', *Proceedings of the National Academy of Sciences*, 105(28), pp. 9697–9702. doi: 10.1073/pnas.0804709105.

Teilum, K., Olsen, J. G. and Kragelund, B. B. (2015) 'Globular and Disordered - the Non-identical Twins in Protein-protein Interactions.', *Frontiers in Molecular Biosciences*, 2, p. 40. doi: 10.3389/fmolb.2015.00040.

Terziyska, N., Mesecke, N., Kozany, C., Neupert, W., Hell, K., Herrmann, J. M. and Baumann, F. (2005) 'A Disulfide Relay System in the Intermembrane Space of Mitochondria that Mediates Protein Import.', *Cell*, 121(7), pp. 1059–1069. doi: 10.1016/j.cell.2005.04.011.

Terziyska, N., Mesecke, N., Kozany, C., Neupert, W., Hell, K., Herrmann, J. M., Lutz, T. and Mokranjac, D. (2005) 'Mia40, a Novel Factor for Protein Import into the Intermembrane Space of Mitochondria is Able to Bind Metal Ions.', *FEBS Letters*, 579(1), pp. 179–184. doi: 10.1016/j.febslet.2004.11.072.

Theurillat, J.-P., Dreier, B., Nagy-Davidescu, G., Seifert, B., Behnke, S., Zürrer-Härdi, U., Ingold, F., Plückthun, A. and Moch, H. (2010) 'Designed Ankyrin Repeat Proteins: a Novel Tool for Testing Epidermal Growth Factor Receptor 2 Expression in Breast Cancer.', *Modern Pathology*, 23(9), pp. 1289–1297. doi: 10.1038/modpathol.2010.103.

Thieulin-Pardo, G., Avilan, L., Kojadinovic, M. and Gontero, B. (2015) 'Fairy "Tails": Flexibility and Function of Intrinsically Disordered Extensions in the Photosynthetic World.', *Frontiers in Molecular Biosciences*, 2, p. 23. doi: 10.3389/fmolb.2015.00023.

Tomasso, M. E., Tarver, M. J., Devarajan, D. and Whitten, S. T. (2016) 'Hydrodynamic Radii of Intrinsically Disordered Proteins Determined from Experimental Polyproline II Propensities.', *PLoS Computational Biology*. Edited by P. Radivojac, 12(1), p. e1004686. doi: 10.1371/journal.pcbi.1004686.

Tompa, P. (2002) 'Intrinsically Unstructured Proteins.', *Trends in Biochemical Sciences*, 27(10), pp. 527–533. doi: 10.1016/S0968-0004(02)02169-2.

Tompa, P. (2003) 'The Functional Benefits of Protein Disorder.', *Journal of Molecular Structure: THEOCHEM*, 666–667, pp. 361–371. doi: 10.1016/j.theochem.2003.08.047.

Tompa, P. (2011) 'Unstructural Biology Coming of Age', *Current Opinion in Structural Biology*, 21(3), pp. 419–425. doi: 10.1016/j.sbi.2011.03.012.

Tompa, P. (2012) 'Intrinsically Disordered Proteins: a 10-year Recap.', *Trends in Biochemical Sciences*, 37(12), pp. 509–516. doi: 10.1016/j.tibs.2012.08.004.

Tompa, P., Bánki, P., Bokor, M., Kamasa, P., Kovács, D., Lasanda, G. and Tompa, K. (2006) 'Protein-Water and Protein-buffer Interactions in the Aqueous Solution of an Intrinsically Unstructured Plant Dehydrin: NMR Intensity and DSC Aspects.', *Biophysical Journal*, 91(6), pp. 2243–2249. doi: 10.1529/biophysj.106.084723.

Tompa, P. and Csermely, P. (2004) 'The Role of Structural Disorder in the Function of RNA and Protein Chaperones.', *FASEB Journal*, 18(11), pp. 1169–75. doi: 10.1096/fj.04-1584rev.

Tompa, P. and Fuxreiter, M. (2008) 'Fuzzy Complexes: Polymorphism and Structural Disorder in Protein-protein Interactions.', *Trends in Biochemical Sciences*, 33(1), pp. 2–8. doi: 10.1016/j.tibs.2007.10.003.

Tompa, P. and Kovacs, D. (2010) 'Intrinsically Disordered Chaperones in Plants and Animals.', *Biochemistry and Cell Biology*, 88(2), pp. 167–74. doi: 10.1139/o09-163.

Tong, Y., Yamaguchi, H., Giaime, E., Boyle, S., Kopan, R., Kelleher, R. J. and Shen, J. (2010) 'Loss of Leucine-rich Repeat Kinase 2 causes Impairment of Protein Degradation Pathways, Accumulation of α-synuclein, and Apoptotic Cell Death in Aged Mice.', *Proceedings of the National Academy of Sciences*, 107(21), pp. 9879–84. doi: 10.1073/pnas.1004676107.

Tran, H. T., Mao, A. and Pappu, R. V (2008) 'Role of Backbone-solvent Interactions in Determining Conformational Equilibria of Intrinsically Disordered Proteins.', *Journal of the American Chemical Society*, 130(23), pp. 7380–7392. doi: 10.1021/ja710446s.

Tripp, K. W. and Barrick, D. (2008) 'Rerouting the Folding Pathway of the Notch Ankyrin Domain by Reshaping the Energy Landscape.', *Journal of the American Chemical Society*, 130(17), pp. 5681–5688. doi: 10.1021/ja0763201.

Trojanowski, J. Q. and Lee, V. M.-Y. (1998) 'Aggregation of Neurofilament and α-synuclein Proteins in Lewy Bodies.', *Archives of Neurology*, 55(2), p. 151. doi: 10.1001/archneur.55.2.151.

Tsytlonok, M., Craig, P. O., Sivertsson, E., Serquera, D., Perrett, S., Best, R. B., Wolynes, P. G. and Itzhaki, L. S. (2013) 'Complex Energy Landscape of a Giant Repeat Protein.', *Structure*, 21(11), pp. 1954–1965. doi: 10.1016/j.str.2013.08.028.

Uversky, V. N. (2010) 'The Mysterious Unfoldome: Structureless, Underappreciated, Yet Vital Part of Any Given Proteome.', *Journal of Biomedicine and Biotechnology*, 2010, pp. 1–14. doi: 10.1155/2010/568068.

Uversky, V. N. (2011a) 'Intrinsically Disordered Proteins from A to Z.', *The International Journal of Biochemistry & Cell Biology*, 43(8), pp. 1090–103. doi: 10.1016/j.biocel.2011.04.001.

Uversky, V. N. (2011b) 'Multitude of Binding Modes Attainable by Intrinsically Disordered Proteins: a Portrait Gallery of Disorder-based Complexes.', *Chemical Society Reviews*, 40(3), pp. 1623–1634. doi: 10.1039/C0CS00057D.

Uversky, V. N. (2017a) 'Intrinsically Disordered Proteins in Overcrowded Milieu: Membrane-less Organelles, Phase Separation, and Intrinsic Disorder.', *Current Opinion in Structural Biology*, 44, pp. 18–30. doi: 10.1016/j.sbi.2016.10.015.

Uversky, V. N. (2017b) 'Protein Intrinsic Disorder-based Liquid-liquid Phase Transitions in Biological Systems: Complex Coacervates and Membrane-less Organelles.', *Advances in Colloid and Interface Science*, 239, pp. 97–114. doi: 10.1016/j.cis.2016.05.012.

Uversky, V. N. and Dunker, A. K. (2010) 'Understanding Protein Non-folding.', *Biochimica et*

Biophysica Acta, 1804(6), pp. 1231–1264. doi: 10.1016/j.bbapap.2010.01.017.

Uversky, V. N., Gillespie, J. R. and Fink, A. L. (2000) 'Why are "Natively Unfolded" Proteins Unstructured under Physiologic Conditions?', *Proteins: Structure, Function, and Genetics*, 41(3), pp. 415–427. doi: 10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.

Uversky, V. N., Oldfield, C. J. and Dunker, A. K. (2005) 'Showing your ID: Intrinsic Disorder as an ID for Recognition, Regulation and Cell Signaling.', *Journal of Molecular Recognition*, 18(5), pp. 343–384. doi: 10.1002/jmr.747.

Uversky, V. N., Oldfield, C. J. and Dunker, A. K. (2008) 'Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept.', *Annual Review of Biophysics*, 37(1), pp. 215–246. doi: 10.1146/annurev.biophys.37.032807.125924.

Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N. and Dunker, A. K. (2007) 'Characterization of Molecular Recognition Features, MoRFs, and Their Binding Partners.', *Journal of Proteome Research*, 6(6), pp. 2351–2366. doi: 10.1021/pr0701411.

Vaquero, A. R., Ferreira, N. E., Omae, S. V, Rodrigues, M. V, Teixeira, S. K., Krieger, J. E. and Pereira, A. C. (2012) 'Using Gene-network Landscape to Dissect Genotype Effects of TCF7L2 Genetic Variant on Diabetes and Cardiovascular Risk.', *Physiological Genomics*, 44(19), pp. 903–914. doi: 10.1152/physiolgenomics.00030.2012.

Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F. and Plückthun, A. (2012) 'Designed Armadillo Repeat Proteins: Library Generation, Characterization and Selection of Peptide Binders with High Specificity.', *Journal of Molecular Biology*, 424(1–2), pp. 68–87. doi: 10.1016/j.jmb.2012.08.029.

Veesler, D., Dreier, B., Blangy, S., Lichiere, J., Tremblay, D., Moineau, S., Spinelli, S., Tegoni, M., Pluckthun, A., Campanacci, V. and Cambillau, C. (2009) 'Crystal Structure and Function of a DARPin Neutralizing Inhibitor of Lactococcal Phage TP901-1: Comparison of DARPin and Camelid VHH Binding Mode.', *Journal of Biological Chemistry*, 284(44), pp. 30718–30726. doi: 10.1074/jbc.M109.037812.

Vekrellis, K., Xilouri, M., Emmanouilidou, E., Rideout, H. J. and Stefanis, L. (2011) 'Pathological Roles of α-synuclein in Neurological Disorders.', *The Lancet Neurology*, 10(11), pp. 1015–1025. doi: 10.1016/S1474-4422(11)70213-7.

Vindigni, A., White, C. E., Komives, E. A. and Di Cera, E. (1997) 'Energetics of

Thrombin−thrombomodulin Interaction.', *Biochemistry*, 36(22), pp. 6674–6681. doi: 10.1021/bi962766a.

Vivès, E. and Lebleu, B. (2003) 'One-pot Labeling and Purification of Peptides and Proteins with Fluorescein Maleimide.', *Tetrahedron Letters*, 44(29), pp. 5389–5391. doi: 10.1016/S0040-4039(03)01304-2.

Vogelstein, B. and Gillespie, D. (1979) 'Preparative and Analytical Purification of DNA from Agarose.', *Proceedings of the National Academy of Sciences*, 76(2), pp. 615–9. doi: 10.1073/pnas.76.2.615.

Walensky, L. D. and Bird, G. H. (2014) 'Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress.', *Journal of Medicinal Chemistry*, 57(15), pp. 6275–6288. doi: 10.1021/jm4011675.

Wales, P., Pinho, R., Lázaro, D. F. and Outeiro, T. F. (2013) 'Limelight on α-synuclein: Pathological and Mechanistic Implications in Neurodegeneration.', *Journal of Parkinson's disease*, 3(4), pp. 415–59. doi: 10.3233/JPD-130216.

Wang, J. T., Smith, J., Chen, B.-C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B. G., Calidas, D., Betzig, E. and Seydoux, G. (2014) 'Regulation of RNA Granule Dynamics by Phosphorylation of Serine-rich, Intrinsically Disordered Proteins in C. elegans.', *eLife*, 3, p. e04591. doi: 10.7554/eLife.04591.

Wang, P., Dong, S., Shieh, J.-H., Peguero, E., Hendrickson, R., Moore, M. A. S. and Danishefsky, S. J. (2013) 'Erythropoietin Derived by Chemical Synthesis.', *Science*, 342(6164), pp. 1357–1360. doi: 10.1126/science.1245095.

Wang, X., Yan, M. H., Fujioka, H., Liu, J., Wilson-Delfosse, A., Chen, S. G., Perry, G., Casadesus, G. and Zhu, X. (2012) 'LRRK2 Regulates Mitochondrial Dynamics and Function through Direct Interaction with DLP1.', *Human Molecular Genetics*, 21(9), pp. 1931–1944. doi: 10.1093/hmg/dds003.

Wang, Z. X. (1995) 'An Exact Mathematical Expression for Describing Competitive Binding of Two Different Ligands to a Protein Molecule.', *FEBS letters*, 360(2), pp. 111–4. doi: 10.1016/0014-5793(95)00062-E.

Weber, P., Ohlendorf, D., Wendoloski, J. and Salemme, F. (1989) 'Structural Origins of High-affinity Biotin Binding to Streptavidin.', *Science*, 243(4887), pp. 85–88. doi: 10.1126/science.2911722.

Wedel, N. and Soll, J. (1998) 'Evolutionary Conserved Light Regulation of Calvin Cycle Activity by NADPH-mediated Reversible Phosphoribulokinase/CP12/Glyceraldehyde-3-phosphate Dehydrogenase Complex Dissociation.', *Proceedings of the National Academy of Sciences*, 95(16), pp.

9699–704. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9689144.

Wedel, N., Soll, J. and Paap, B. K. (1997) 'CP12 Provides a New Mode of Light Regulation of Calvin Cycle Activity in Higher Plants.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(19), pp. 10479–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9294236.

Werbeck, N. D. and Itzhaki, L. S. (2007) 'Probing a Moving Target with a Plastic Unfolding Intermediate of an Ankyrin-repeat Protein.', *Proceedings of the National Academy of Sciences*, 104(19), pp. 7863–7868. doi: 10.1073/pnas.0610315104.

Werbeck, N. D., Rowling, P. J. E., Chellamuthu, V. R. and Itzhaki, L. S. (2008) 'Shifting Transition States in the Unfolding of a Large Ankyrin Repeat Protein.', *Proceedings of the National Academy of Sciences*, 105(29), pp. 9982–9987. doi: 10.1073/pnas.0705300105.

West, A. B., Cowell, R. M., Daher, J. P. L., Moehle, M. S., Hinkle, K. M., Melrose, H. L., Standaert, D. G. and Volpicelli-Daley, L. A. (2014) 'Differential LRRK2 Expression in the Cortex, Striatum, and Substantia Nigra in Transgenic and Nontransgenic Rodents.', *Journal of Comparative Neurology*, 522(11), pp. 2465–2480. doi: 10.1002/cne.23583.

Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. and Plückthun, A. (2008) 'Folding and Unfolding Mechanism of Highly Stable Full-consensus Ankyrin Repeat Proteins.', *Journal of Molecular Biology*, 376(1), pp. 241–257. doi: 10.1016/j.jmb.2007.11.046.

Willert, K. and Jones, K. A. (2006) 'Wnt Signaling: is the Party in the Nucleus?', *Genes & Development*, 20(11), pp. 1394–404. doi: 10.1101/gad.1424006.

Winner, B., Melrose, H. L., Zhao, C., Hinkle, K. M., Yue, M., Kent, C., Braithwaite, A. T., Ogholikhan, S., Aigner, R. and Winkler, J. (2011) 'Adult Neurogenesis and Neurite Outgrowth are Impaired in LRRK2 G2019S Mice.', *Neurobiology of Disease*, 41(3), pp. 706–716. doi: 10.1016/j.nbd.2010.12.008.

Wong, C. K., Luo, W., Deng, Y., Zou, H., Ye, Z. and Lin, S.-C. (2004) 'The DIX Domain Protein Coiled-coil-DIX1 Inhibits c-Jun N-terminal Kinase Activation by Axin and Dishevelled through Distinct Mechanisms.', *Journal of Biological Chemistry*, 279(38), pp. 39366–39373. doi: 10.1074/jbc.M404598200.

Wong, H.-C., Bourdelas, A., Krauss, A., Lee, H.-J., Shao, Y., Wu, D., Mlodzik, M., Shi, D.-L. and Zheng, J. (2003) 'Direct Binding of the PDZ Domain of Dishevelled to a Conserved Internal Sequence in the C-terminal Region of Frizzled.', *Molecular Cell*, 12(5), pp. 1251–1260. doi: 10.1016/S1097-

2765(03)00427-1.

Wood, W. B. (1966) 'Host Specificity of DNA Produced by Escherichia coli: Bacterial Mutations Affecting the Restriction and Modification of DNA', *Journal of Molecular Biology*, 16(1), p. 118–IN3. doi: 10.1016/S0022-2836(66)80267-X.

Worrell, B. T., Malik, J. A. and Fokin, V. V (2013) 'Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-catalyzed Azide-alkyne Cycloadditions', *Science*, 340(6131), pp. 457–460. doi: 10.1126/science.1229506.

Wright, P. E. and Dyson, H. J. (1999) 'Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-function Paradigm.', *Journal of Molecular Biology*, 293(2), pp. 321–331. doi: 10.1006/jmbi.1999.3110.

Wright, P. E. and Dyson, H. J. (2009) 'Linking Folding and Binding', *Current Opinion in Structural Biology*, 19(1), pp. 31–38. doi: 10.1016/j.sbi.2008.12.003.

Wu, G., Xu, G., Schulman, B. A., Jeffrey, P. D., Harper, J. W. and Pavletich, N. P. (2003) 'Structure of a β-TrCP1-Skp1-β-catenin Complex.', *Molecular Cell*, 11(6), pp. 1445–1456. doi: 10.1016/S1097-2765(03)00234-X.

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N. and Obradovic, Z. (2007a) 'Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions.', *Journal of Proteome Research*, 6(5), pp. 1882–1898. doi: 10.1021/pr060392u.

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N. and Obradovic, Z. (2007b) 'Functional Anthology of Intrinsic Disorder. 2. Cellular Components, Domains, Technical Terms, Developmental Processes, and Coding Sequence Diversities Correlated with Long Disordered Regions.', *Journal of Proteome Research*, 6(5), pp. 1899–1916. doi: 10.1021/pr060393m.

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N. and Obradovic, Z. (2007c) 'Functional Anthology of Intrinsic Disorder. 3. Ligands, Post-Translational Modifications, and Diseases Associated with Intrinsically Disordered Proteins.', *Journal of Proteome Research*, 6(5), pp. 1917–1932. doi: 10.1021/pr060394e.

Xing, Y., Clements, W. K., Kimelman, D. and Xu, W. (2003) 'Crystal Structure of a β-catenin/Axin Complex Suggests a Mechanism for the β-catenin Destruction Complex.', *Genes & Development*, 17(22), pp. 2753–64. doi: 10.1101/gad.1142603.

Xing, Y., Clements, W. K., Le Trong, I., Hinds, T. R., Stenkamp, R., Kimelman, D. and Xu, W. (2004) 'Crystal Structure of a β-catenin/APC Complex Reveals a Critical Role for APC Phosphorylation in APC Function.', *Molecular Cell*, 15(4), pp. 523–33. doi: 10.1016/j.molcel.2004.08.001.

Xing, Y., Takemaru, K.-I., Liu, J., Berndt, J. D., Zheng, J. J., Moon, R. T. and Xu, W. (2008) 'Crystal Structure of a Full-length β-catenin.', *Structure*, 16(3), pp. 478–87. doi: 10.1016/j.str.2007.12.021.

Xu, J., Kao, S.-Y., Lee, F. J. S., Song, W., Jin, L.-W. and Yankner, B. A. (2002) 'Dopamine-dependent Neurotoxicity of α-synuclein: a Mechanism for Selective Neurodegeneration in Parkinson Disease.', *Nature Medicine*, 8(6), pp. 600–6. doi: 10.1038/nm0602-600.

Xu, W. and Kimelman, D. (2007) 'Mechanistic Insights from Structural Studies of β-catenin and its Binding Partners.', *Journal of Cell Science*, 120(Pt 19), pp. 3337–44. doi: 10.1242/jcs.013771.

Xu, W., Lau, Y. H., Fischer, G., Tan, Y. S., Chattopadhyay, A., de la Roche, M., Hyvönen, M., Verma, C., Spring, D. R. and Itzhaki, L. S. (2017) 'Macrocyclized Extended Peptides: Inhibiting the Substrate-recognition Domain of Tankyrase.', *Journal of the American Chemical Society*, 139(6), pp. 2245–2256. doi: 10.1021/jacs.6b10234.

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. and Uversky, V. N. (2010) 'PONDR-FIT: A Meta-predictor of Intrinsically Disordered Amino Acids.', *Biochimica et Biophysica Acta*, 1804(4), pp. 996–1010. doi: 10.1016/j.bbapap.2010.01.011.

Xue, B., Romero, P. R., Noutsou, M., Maurice, M. M., Rüdiger, S. G. D., William, A. M., Mizianty, M. J., Kurgan, L., Uversky, V. N. and Dunker, A. K. (2013) 'Stochastic Machines as a Colocalization Mechanism for Scaffold Protein Function.', *FEBS Letters*, 587(11), pp. 1587–91. doi: 10.1016/j.febslet.2013.04.006.

Yamada, J., Phillips, J. L., Patel, S., Goldfien, G., Calestagne-Morelli, A., Huang, H., Reza, R., Acheson, J., Krishnan, V. V, Newsam, S., Gopinathan, A., Lau, E. Y., Colvin, M. E., Uversky, V. N. and Rexach, M. F. (2010) 'A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins.', *Molecular & Cellular Proteomics*, 9(10), pp. 2205–2224. doi: 10.1074/mcp.M000035-MCP201.

Yamada, T., Takaoka, A. S., Naishiro, Y., Hayashi, R., Maruyama, K., Maesawa, C., Ochiai, A. and Hirohashi, S. (2000) 'Transactivation of the Multidrug Resistance 1 Gene by T-cell Factor 4/β-catenin Complex in Early Colorectal Carcinogenesis.', *Cancer Research*, 60(17), pp. 4761–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10987283.

Yamamoto, H., Komekado, H. and Kikuchi, A. (2006) 'Caveolin Is Necessary for Wnt-3a-dependent Internalization of LRP6 and Accumulation of β-catenin.', *Developmental Cell*, 11(2), pp. 213–223. doi: 10.1016/j.devcel.2006.07.003.

Yang, J., Zhang, W., Evans, P. M., Chen, X., He, X. and Liu, C. (2006) 'Adenomatous Polyposis Coli (APC) Differentially Regulates β-catenin Phosphorylation and Ubiquitination in Colon Cancer Cells.', *The Journal of Biological Chemistry*, 281(26), pp. 17751–7. doi: 10.1074/jbc.M600831200.

Yang, W., Gelles, J. and Musser, S. M. (2004) 'Imaging of Single-molecule Translocation through Nuclear Pore Complexes.', *Proceedings of the National Academy of Sciences*, 101(35), pp. 12887–12892. doi: 10.1073/pnas.0403675101.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) 'Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors.', *Gene*, 33(1), pp. 103–19. doi: 10.1016/0378-1119(85)90120-9.

Yin, H., Kauffman, K. J. and Anderson, D. G. (2017) 'Delivery Technologies for Genome Editing.', *Nature Reviews Drug Discovery*, 16(6), pp. 387–399. doi: 10.1038/nrd.2016.280.

Yoshida, R., Kimura, N., Harada, Y. and Ohuchi, N. (2001) 'The Loss of E-cadherin, α- and β-catenin Expression is Associated with Metastasis and Poor Prognosis in Invasive Breast Cancer.', *International Journal of Oncology*, 18(3), pp. 513–20. doi: 10.3892/ijo.18.3.513.

Yuan, J. and Yankner, B. A. (2000) 'Apoptosis in the Nervous System.', *Nature*, 407(6805), pp. 802–9. doi: 10.1038/35037739.

Yue, Z. and Lachenmayer, M. L. (2011) 'Genetic LRRK2 Models of Parkinson's Disease: Dissecting the Pathogenic Pathway and Exploring Clinical Applications.', *Movement Disorders*, 26(8), pp. 1386–97. doi: 10.1002/mds.23737.

Yun, H. J., Park, J., Ho, D. H., Kim, H., Kim, C.-H., Oh, H., Ga, I., Seo, H., Chang, S., Son, I. and Seol, W. (2013) 'LRRK2 Phosphorylates Snapin and Inhibits Interaction of Snapin with SNAP-25.', *Experimental & Molecular Medicine*, 45(8), p. e36. doi: 10.1038/emm.2013.68.

Zeng, X., Huang, H., Tamai, K., Zhang, X., Harada, Y., Yokota, C., Almeida, K., Wang, J., Doble, B., Woodgett, J., Wynshaw-Boris, A., Hsieh, J.-C. and He, X. (2008) 'Initiation of Wnt Signaling: Control of Wnt Coreceptor LRP6 Phosphorylation/Activation via Frizzled, Dishevelled and Axin Functions.', *Development*, 135(2), pp. 367–75. doi: 10.1242/dev.013540.

Zeng, X., Tamai, K., Doble, B., Li, S., Huang, H., Habas, R., Okamura, H., Woodgett, J. and He, X. (2005)

'A Dual-kinase Mechanism for Wnt Co-receptor Phosphorylation and Activation.', *Nature*, 438(7069), pp. 873–877. doi: 10.1038/nature04185.

Zhang, C., Bao, W., Rong, Y., Yang, H., Bowers, K., Yeung, E. and Kiely, M. (2013) 'Genetic Variants and the Risk of Gestational Diabetes Mellitus: a Systematic Review.', *Human Reproduction Update*, 19(4), pp. 376–390. doi: 10.1093/humupd/dmt013.

Zhou, Y., Simpson, S., Charlesworth, J. C., van der Mei, I., Lucas, R. M., Ponsonby, A.-L. and Taylor, B. V (2017) 'Variation within MBP Gene Predicts Disease Course in Multiple Sclerosis.', *Brain and Behavior*, 7(4), p. e00670. doi: 10.1002/brb3.670.